Multilingualism Doctoral School

University of Pannonia

A MIXED-METHODS STUDY ON VERBAL ABILITIES AND COGNITIVE FLEXIBILITY OF HUNGARIAN LEARNERS IN CLIL AND GENERAL LANGUAGE PROGRAMMES

DOI:10.18136/PE.2021.791

PhD Thesis

by

Ágnes Sántha-Malomsoki

Supervisors:

Prof. Dr. Valéria Csépe | Prof. Dr. Judit Navracsics

Veszprém

2021

This dissertation, written under the direction of the candidate's committee and approved by the members of the committee, has been presented to and accepted by the Faculty of Modern Philology and Social Sciences in partial fulfilment of the requirements for the degree of Doctor of Philosophy. The content and research methodologies presented in this work represent the work of the candidate alone.

Ágnes Sántha-Malomsoki June, 2021

Candidate ____________ Date

Dissertation Committee:

__________________________________ June, 2021

Chairperson ____________ Date

__________________________________ June, 2021

First Reader ____________ Date

__________________________________ June, 2021

Second Reader ____________ Date

A MIXED-METHODS STUDY ON VERBAL ABILITIES AND COGNITIVE FLEXIBILITY OF HUNGARIAN LEARNERS IN CLIL AND GENERAL LANGUAGE PROGRAMMES

Thesis for obtaining a PhD degree in the Doctoral School of Multilingualism of the University of Pannonia

in the branch of Modern Philology and Social Sciences

Written by Ágnes Sántha-Malomsoki

Supervisors:

Prof. Dr. Valéria Csépe   Prof. Dr. Judit Navracsics

propose acceptance:

_______________ (yes / no)   _______________ (yes / no)

______________________________   ______________________________

(Supervisor's signature)   (Supervisor's signature)

As a reviewer, I propose acceptance of thesis:

Name of Reviewer: _____________ (yes / no)

______________________________ (Reviewer's signature)

The PhD-candidate has achieved _______ (%) a the public discussion,

Veszprém, __________________ (Date)

____

(Chairman of the Committee's signature)

The grade of the PhD Diploma _______ (%)

Veszprém, __________________ (Date)

____

(Chairman of UDHC's signature)

Dissertation Abstract

The number of schools offering CLIL (Content and Language Integrated Learning) programmes is increasing in Hungary. They consider language development as a natural and dynamic process in which learners play an active role. These programmes are characterized by the parallel use of both languages with the general aim of supporting conceptual knowledge construction in either language. Programmes like these provide intensive exposure to authentic second language embedded in meaningful practices with the final aim of making learners achieve an officially declared language level. This different L2 teaching approach can cause learners' qualitatively different levels of knowledge, learning paths and mental sets. For this reason, in this study, we applied Mixed Methods to investigate whether extensive (CLIL) and general second language use among instructed conditions result in different verbal and cognitive outcomes that are detectable via either quantitative and qualitative methods.

The study was designed in accordance with Creswell's Sequential Explanatory Design (2012) in which a quantitative large sample study is followed by a qualitative small-sample study. In the model the qualitative method serves as the main method. The data received this way further refine the results to serve deeper understanding. In the first phase of the research an experimental group (CLIL group, N=69) and a control group (N=73) were compared with the involvement of a language experience and proficiency questionnaire (LEAP-Q), a selective attention test (d2-R) and a phonemic fluency test (in the first language (L1=Hungarian) and the second language (L2=English). LEAP-Q questionnaire provided background information about learners' attitude, exposure and assumed level regarding the L1 and L2 in both groups. D2-R test was chosen to explore whether there is a difference between the two groups in terms of selective attention that is usually cited to be more enhanced in bilinguals as a result of constant shifting between the L1 and L2. The purpose of the application of the phonemic fluency tests was twofold. Firstly, these test types provide information about learners' executive abilities and second, they might also refer to the size of their mental lexicon. Executive functioning was measured by variables of shifting and clustering while the size of the mental lexicon was defined by the total number of generated words and that of words from different word classes. Since word retrieval is often cited to be slower for bilinguals in comparison to monolinguals in the scientific literature, we expected results accordingly.

For this reason, in the first phase a large-scale quantitative data collection and analysis was carried out with the aim of exploring specific verbal and cognitive patterns in the test outcomes. As a result, four different groups have been defined: a CLIL 'high' (N=3), a control 'high' (N=3), a CLIL 'low' (N=3) and a control 'low' (N=3) group. Those learners have been selected for the 'high' groups who achieved exceptionally high results in all test types compared to their group results. Conversely, 'low' group learners achieved the lowest results in all test types. We assumed that superiority in the tests would be reflected in the way learners form their opinions on L2-related questions. To gain insight in learners' thinking patterns, a structured interview served as a tool in the second phase of the research.

The test outcomes revealed no significant difference related to selective attention; however, significant differences have been found for most of the variables related to phonemic fluency in the L2, indicating higher level of executive functioning in case of the experimental group. Findings of the qualitative interview analyses are in line with these test outcomes in case of the CLIL 'high' group. Therefore, our final conclusion is that extensive second language use paired with CLIL methodology might contribute to strategy use not only in tasks of lexical retrieval but in an interview situation as well, when the flow of ideas are needed.

To Mum and Dad

Acknowledgements

The process of writing a dissertation is not only a challenge of intellect but that of self-journey. In the following lines, I would like to express my sincere gratitude to all those people who supported me during this long journey. Without their encouragement and guidance, I would not have succeeded.

I am genuinely grateful for having had the support and guidance of two such professional excellences as Prof. Dr. Valéria Csépe and Prof. Dr. Judit Navracsics. Their commitment to the highest standards was inspiring and motivating. Working with them is a life experience I will always remember.

I owe a dept of thanks to all the professors at Multilingualism Doctoral School who helped me see the world beyond the visible spectrum. I truly appreciate the opportunity of learning from all of them.

Special thanks go to all my family members, who swung me through the difficult times. I am extremely thankful to Dr. Kálmán Sántha, for being a caring and loving husband and also for sharing his expertise on qualitative research methodology with me anytime I asked for it. I also want to thank my son, Áron, for his understanding and empowering love. I must extend my heartfelt gratitude to my parents and brother who set good examples by making me persistent and obstinate enough to struggle hard for my goals.

I would like to thank my colleagues, Ágnes Zita Kauthné Szabó and Tamás Barta for constantly encouraging me to move ahead. I feel fortunate to work with them and I highly appreciate their humanity, engagement and professionalism that characterize their personality. The accomplishment of this dissertation could not have been possible without Dr. László Pribék's valuable suggestions on the applied statistical methods.

I am thankful to all my former and present colleagues, and the students at II. Rákóczi Ferenc Hungarian-English Bilingual School with whom I worked until 2019. I was very lucky to be a part of a school with such high standards. This experience left a mark on my heart forever.

At last, but not least I am infinitely grateful to the parents who gave their consent to their children's participation in the research. Without their openness this dissertation would have remained just an idea.

List of Tables

List of Figures

TABLE OF CONTENTS

CHAPTER 1. Introduction

Today, in this dynamically changing world, new competences such as adaptability, flexible thinking, co-operation, problem solving, and communication (operations all related to ‘hot’ and ‘cold’ aspects of the executive functions) are increasingly appreciated by the labour market. These personal attitudes are more valued than certificates and credentials of knowledge, since they cannot be learnt in the way as book-based information at school. If educators intend to teach in line with future needs, they must focus more on learners’ personality development than on their subject knowledge. Consequently, traditional ‘old-school’ teaching and learning methods cannot support the emergence and maintenance of that proactive attitude that goes beyond the passive acceptance of ready-made thoughts. However, educational practices entrenched over the centuries are difficult to override and the efforts made to reform them are in their infancy. Reinterpreting the role of the teacher, the scope of teaching and integrating innovative practices in the process are the initial steps of this paradigm shift.

However, some forerunners have already been operating educational practices that seem to reflect this novel approach. One such example for these programmes is Content and Language Integrated Learning (hereinafter CLIL) that is seen as a dual-focused educational programme favoured worldwide and characterized by fundamentally different pedagogical and educational practices in comparison to mainstream second language (hereinafter L2) programmes. Knowledge in CLIL programmes emerges from the meaningful use of language and represents a qualitatively different educational approach. Thus, these programmes require and support the integration of novel subject methods and promote the acceptance of interchangeability of teacher-learner roles in the school context.

Conducting research on CLIL is rooted in my sixteen years of experience as an L2 language and subject teacher. I have taught learners of all age groups, from first graders to adults. I have had experience with primary school learners at different levels with different cognitive abilities, language backgrounds and specifications (learners studying according to a general, English-specified or CLIL curriculum). Although the difference among these classes was overt in many aspects (attitude towards the L2, motivation in learning, curiosity about new information, flow of ideas, approaching questions from unusual perspectives and coping with task-related problems), I regularly experienced my colleagues' scepticism on the efficiency of CLIL programmes. They often listed a vast number of reasons why extensive L2 teaching is not necessary at an early age. Their concerns mostly revolved around one particular topic: the perfect age for initiating intensive L2 learning. Most of these colleagues were teaching reading, writing and maths for lower graders and were not officially educated in L2 development, the L2 and CLIL methodology. Nevertheless, they often argued that such intensive learning of a second language hinders the development of the mother tongue in many ways. It slows down reading and causes problems in spelling in the L1. They were terrified by the emergence of the L2 in the L1 class since they were unaware of the constant interaction between the languages. Despite my persistent protest, they determinedly insisted on the misconception that if learners did not master the L1, it was unnecessary to teach the L2 with such intensity. They did not even consider that language development in general, does not have an endpoint, thus waiting for the perfect timing regarding the initiation of the L2 does not make sense, since it might be different for learners.

Issues like these led me to address the subject in depth. The following research was highly inspired by my pedagogical urge to reveal that knowledge gained via CLIL is far beyond a complex intermediate language exam that is generally set as a primary goal to be achieved by the end of the programme.

1.1. Background and rationale of the study

Parallel to the emerging demand for a paradigm shift in education, English is increasingly becoming a basic skill. Today, the proliferation of info communication technologies provides many ways to access information in any languages, thus teachers are no longer the primary sources of knowledge. As a result, the process of learning is no longer as transparent and linear as it once was. Consequently, teachers' status in this process has also been transforming and they need to be open to new pedagogical practices (Prievara, 2015; Selwyn, 2017; Váczi, 2018).

This kind of novelty has long been present in bilingual programmes. These programmes, in the broadest sense, are exceptionally popular worldwide, though they are also very dissimilar in many aspects, e.g.: languages, through which education is implemented; programme structures (lasting from a few months to years) and contents (what school subjects are chosen); final objectives; and social contexts. Moreover, the implementation of CLIL programmes is mostly adjusted to the core curricula of the different nations. It is also not rare that some programmes focus rather on transferring content than language (Goris, 2019). The only feature they commonly share is the extensive use of an L2 supported by a methodology which is different from that of general programmes (Mehisto et al, 2009; Ball et al, 2015; Van Mensel et al., 2019).

CLIL programmes in Europe, in the strict sense of the term, have always been accused of being elitist and excluding underprivileged individuals. In the Netherlands, learners can be enrolled to the programs only with taking part in a selection process whose application is officially permitted for schools. As a result, children from more privileged families have been enrolled to these programmes, thus their learning outcomes might differ from those learning in mainstream programmes. In Spain, CLIL contents and policy may vary from school to school. In Sweden or Germany, CLIL is typically self-selective, if learners who enrol the programme are more motivated and have a higher scholastic aptitude. In Belgium and Hungary learners' enrolment for the programme is strictly rule-governed (Mihály, 2009; Mehisto et al., 2009; Dallinger, et al., 2016; Van Mensel, et al., 2019; Goris, 2019; Escobar Urmenta, 2019; de Boer & Leontjev, 2020). In Hungarian state schools, first graders cannot be tested for their skills prior to entering school; therefore these programmmes are officially available to anyone regardless of social or financial status. Learners primarily must be admitted to schools in their own districts. Since in Hungary more and more parents are considering the potentials of well-founded L2 knowledge, many schools have been launching CLIL programmes. This way they can ensure future stability for the institution and staff.

Despite the popularity of the programme, studies conducted on it are not only rare but also unreliable to some extent as they apply various research methods and involve participants with different backgrounds. Most studies on CLIL conducted in Europe, approached it from the perspective of language pedagogy and focused on the positive language outcomes. However, results might lead to false conclusions if the framework, in which CLIL is implemented, is ignored. If the practice of a certain skill is more pronounced in one program, gaining better research outcomes in it is not surprising compared to a different program that develops all skills equally. As in recent studies these frames are rarely detailed, and teachers' methodological practices are often unknown, positive results can suggest different interpretations. The lack of studies on CLIL programmes can be traced back to the diverse political, educational, and financial conditions in the European countries, which makes comparability problematic. If studies are conducted, participants are not controlled for factors such as learners' socio-economic status, learners' abilities, or the quality of instruction (Dallinger, et al., 2016).

In the past twenty years, there have been two waves in CLIL research. Early studies had a cross-sectional design and focused primarily on language outcomes (Dallinger, et al., 2016). Most of them have been conducted in Spain, though results of investigations in Austria, Germany, Holland and Italy have also contributed to the research in this field (Pérez-Cañado, 2012; Pérez-Cañado, 2018; Goris, 2019). In Pérez-Cañado's (2012) meta-analysis of early research on the European CLIL practice, a vast array of language-related positive outcomes (vocabulary size, receptive skills, fluency in speaking and writing, lexical and syntactic complexity, creativity and risk-taking) are listed. However, many of them contain inconsistent findings due to the mismatch between school programmes or participants' languages or family backgrounds. Moreover, hardly any investigates contributions that can be unequivocally related to CLIL methodology (Goris, 2019). Critical voices often argue that there is no matching between the experimental and control groups in terms of learner abilities, language level and scholastic aptitude; thus research outcomes might fail to reflect reality. Moreover, learners' socioeconomic status or the initial conditions in terms of L2 learning might also be crucial factors that further research has to take into consideration (Pérez-Cañado, 2012; Verspoor, et al., 2015; Dallinger, et al., 2016; Goris, 2019). Recent studies on CLIL are mainly longitudinal in their designs and cover the mentioned countries. They focus on secondary school learners' L2 skills, the impact of CLIL on L1 and L2 and learners' motivation (Pérez-Cañado, 2018; Goris, 2019). However, the results of these longitudinal studies often report both on significant differences and no differences in comparison to non-CLIL participants (Dallinger, et al., 2016).

If we do not focus on language pedagogical aspects of extensive L2 use but approach it from linguistic or cognitive perspectives, research outcomes are also not very consistent. In many studies, bilingual children are reported to underperform their monolingual counterparts. Their poor results on lexical retrieval (related to the semantic memory) are explained by various reasons like the number of competing languages and words, the joint activation of the languages, the less time recruited to either language or the many other individual factors that might have influence on test outcomes (Abutalebi & Green, 2007; Bialystok et al., 2010; Bialystok & Poarch, 2014; Sullivan et al., 2017). However, the opposite is observed in case of specific executive functions (Costa, et al., 2009; Bialystok et al., 2010; Luk et al., 2011; de Groot & Dukes, 2011; Escobar, 2018). Bilingual children's better performance is often assumed to be the indicator of their advanced executive functions even if their testability is not a clear case. Studies on bilinguals' executive functions mainly involve tests that aim to investigate working memory, inhibitory control and shifting abilities separately, although they are considered to operate in an interrelated manner.

Given that research on CLIL has mainly investigated the programme from the aspect of L2 pedagogy, I aimed at approaching it from a different perspective: bilingualism. For this reason, I applied a mixed methodology research, in which a large-scale statistical analysis is followed by qualitative content analysis. The reason for applying Mixed Methods in this study is twofold: first, I aimed to draw attention to learners' individual developmental trajectories from which specific patterns can be revealed by means of quantitative data analysis, and secondly, I assumed that these patterns can also be captured via qualitative content analysis in which the deep layers of learners' thinking are mapped.

To my knowledge this is the first study that aims at comparing executive functioning and verbal abilities of Hungarian CLIL learners and traditional learners of English using Mixed Methods.

1.2 Bilingualism and bilingual education

Even though there is no exact number of people using more than one language on a daily basis, it is estimated that half of the world's population falls into the 'bilingual' category (Grosjean, 2013). The reasons for this might be various: there are many countries in which more than one official or local language is used; certain positions require the use of multiple languages and trading worldwide is unfeasible without language. Education is also highly involved in the spread of second languages (Jessner, 2008; Grosjean, 2013; Cook&Singleton, 2014).

Bilingualism is a multidisciplinary concept with many definitions that belong to different disciplines and hence among which the boundaries are blurred. In the general sense of the term a bilingual person in contrast with a monolingual can speak two languages. Some of the difficulties of giving an exact definition might arise from the fact that many kinds and degrees of bilingualism and bilingual situations exist, and the phenomenon is often viewed from a monolingual perspective. Considering bilingualism a static condition, early approaches described it as a dichotomous concept and observed it from a linguistic perspective with focus on bilinguals' level of proficiency. According to Bloomfield's (1933) maximalist view, a bilingual can use both languages perfectly. As Navracsics (2010) points it out even a monolingual does not know their mother tongue at a hundred per cent: the registers that are owned by different social groups such as slang words or profession-related terminology are often context and time-dependent. Grosjean (2013; 2016) in his Complementary Principle (CP) claims that bilinguals' languages are normally used in different domains (e.g. home or school) with different people (e.g. family and teachers) and frequency, in various modalities (visual and auditory), and this has an impact on the level of their fluency (Lesznyák, 1996; Navracsics, 2007a; Grosjean, 2013; Cook & Singleton, 2014). Grosjean (2013) adds that even if they have different proficiency levels in the four language skills of their two languages, they can use their languages to achieve their goals. His approach is defined as functional bilingualism. Proficiency, fluency, frequency, and the number of domains in which the languages are used are strongly correlated: the more situations a language can be used in, the higher level of mastery can be achieved (Grosjean, 2016). Since language use is often linked to certain domains, in certain situations instead of L1, L2 can be the dominant language (Grosjean, 2013, 2016). If the relation of languages to one another is considered, coordinate, compound and subordinate bilingualism can be distinguished (Weinreich, 1953). In case of coordinate bilinguals, the languages are not related at the conceptual level, hence kept apart. Conversely, compound bilinguals' languages are connected by the same concept. Subordinate bilinguals relate and understand the L2 through their L1 (Navracsics, 2007a; Navracsics, 2008; Cook & Singleton, 2014). Another aspect from which bilingualism can also be approached is the chronological order of the acquisition of languages. In simultaneous bilingual language acquisition both languages are present from birth in the child's life. Navracsics (2008) claims that bilingualism is not a fixed, stable state and even a slight change in the person's environment might cause the dominance of one language over the other resulting in an uneven development. Successive language acquisition, by contrast, focuses on second language acquisition that might have an onset any time across the lifespan. Bilingualism was also viewed from the age of language acquisition: whether it starts before or after the critical period. However, the existence of the critical period for humans is hotly debated (c.f. Navracsics, 2008). If the environment, in which L2 acquisition is taking place, is under scrutiny, informal/naturalistic or formal/instructed settings can be distinguished. As opposed to informal/naturalistic learning, formal/instructed learning is implemented in schools in which the programmes are adjusted to the learners' needs.

In the first half of the last century, bilingualism was regarded as an obstacle, a kind of mental retardation with which bilinguals need to cope. This conception, however, reflected a monolingual perspective from which bilingualism is different, indeed (Grosjean, 1989; Cook & Singleton, 2014). Bilingualism differs from monolingualism in at least two ways: quantitatively (the number of used languages) and qualitatively. Considering the latter one it can be claimed that bilinguals are multicompetent language users. They are also different from monolinguals in terms of the level of language awareness that is the ability to intentionally switch focus on form, function and meaning. In case of bilinguals, it is labelled as metalinguistic awareness (Jessner, 2008; Cook & Singleton, 2014). The development of metalinguistic awareness is often escorted and followed by that of divergent thinking, interactional competence, communicative sensitivity, and translation skills (Jessner, 2008). Moreover, the active use of a second language has an impact on the individual's emotional life and the use of their first language as well (Cook & Singleton, 2014).

According to Grosjean's (2013) psycholinguistic approach, bilinguals can adjust to different environmental changes in their language behaviour, i.e., function in different modes depending on the people they are communicating to, the topic they are dealing with or the situation they are being involved in, and many more. With this 'mode-concept' he questions the verisimilitude of language non-selectiveness. He also claims that monolingual and bilingual modes are towards the two imagined end points of the continuum, and it is not scarce that bilinguals switch their modes many times during a day. Even if one of the languages is in use (base language) and the other is less active (rather than inhibited), bilinguals use their languages differently with bilinguals and monolinguals. The less active language is often brought in the discourse if the chat partner is the speaker of the same languages (Grosjean, 2013). However, Yu and Schwieter (2018) claim that it is not only the context that has an impact on the activation of the languages, but also the interplay of different individual characteristics, such as language proficiency and dominance. They also suggest the consideration of language mode (instead of controlling for it) in research projects to avoid the alteration of the bilingual experience that might result in different outcomes.

1.2.1 The earlier, the better?

Different theories have come up to explain age-related individual differences in language learning. The 'Critical Period Hypothesis' posited that children are more effective language learners than adults due to a natural inborn age constraint for the initiation of language learning. Lenneberg (1967) argued that (first and second) language acquisition takes place effectively only from early childhood to puberty, after which brain lateralization is complete hence language acquisition is less successful. However, this hypothesis was refined to 'sensitive period' since many related questions remained unanswered. The term 'sensitive period' refers to the less strict boundaries of the period and the huge individual differences among language users. It suggests that there are multiple periods in which learners are more attuned to different aspects of language (phonology, morphology, syntax) (Saville-Troike & Barto, 2012). Cook and Singleton (2014) assume that learning a second language might be supported by different mechanisms in comparison to L1 learning. For this reason, it is not typical for learners to achieve a native-like proficiency, though it is not impossible or rare. Furthermore, in L2 learning in instructed conditions, direct attention is unlikely given to pronunciation. However, Navracsics (2008) emphasizes the remarkable role of pronunciation in L2 use, referring to the ignorance shown by native speakers for conversations in which they can hardly focus on the content if the chat partner's pronunciation is inappropriate. For this reason Nikolov (2011) suggests the consciuous involvement of authentic materials in the L2 classes with which more focus can be given to pronunciation and intonation.

The two frequently cited approaches regarding the developmental constraints of second language acquisition are related to the view on access to the Chomskyan Universal Grammar and the DeKeyser's implicit versus explicit learning mechanisms. In the Chomskyan framework, four possibilities regarding the access of UG are differentiated:

no access, full access, partial access, or indirect access (Cook & Singleton, 2014). On the contrary, DeKeyser (2000) claims that developmental constraints have a role only in implicit learning mechanisms, when non-conscious learning takes place. Navracsics (2008) confirms that the procedural memory, is the underlying system for implicit learning. Consequently, in case of late L2 beginners processing the L2 is supported by the declarative memory. Although frequent L2 use triggers implicit learning and when the individual is able to use both languages, the memory system becomes similar to that of a monolingual's. This is assumed to be the reason why late starters can produce a faster rate of learning, though it is not scientifically evidenced yet (Pfenninger & Singleton, 2019).

In addition, many different factors might influence learning outcomes among instructed settings. Motivation and perseverance are those two crucial factors that support language learning at any age (Navracsics, 2008; Navracsics, 2010; Cook & Singleton, 2014; Griffiths, 2018). In their comprehensive review on this field, Pfenninger and Singleton (2019) conclude that older learners have a high instrumental (goal-oriented) motivation in comparison to young learners, whose intrinsic motivation is mainly influenced by the quality of teaching. Therefore, starting learning an L2 early in itself does not guarantee the maintenance of motivation in the long run, but attitude and beliefs on language learning do so. Furthermore, there are many other factors like socio-affective factors, the process of myelination of the nerves, cognitive variables, role of literacy skills, cognitive style, gender, personality, aptitude and even learning strategy that might also contribute to the fact that older learners participating in traditional language lessons seem less proficient in the long run compared to early starters (Singleton & Pfenninger, 2018; Griffiths, 2018, Saville-Troike & Barto 2012; Pfenninger & Singleton, 2019). However, age-related findings on the advantages of an early start as opposed to the later one are at least mixed.

1.2.2 Second language development – the Dynamic Systems Theory (DST) approach

Parallel to first language acquisition, a range of underlying theoretical approaches including behaviourism, structuralism, cognitivism, socio-culturalism, and humanism framed second language teaching approaches in the past, each of them emphasizing a different aspect of language acquisition (linguistic, psychological, and social).

One of the recent constructionist theories on language development adopted the approach used by physical and natural sciences for touching on the dynamism of many-folded variables. The Dynamic Systems Theory (DST) states that first and second language developments are not linear and static processes with end points, but rather dynamic ones (de Bot et al., 2012) referring to the several interconnected variables that change (grow and decline) and interact over time (Jessner, 2008). This theory also supports that L2 acquisition is driven by cognitive abilities and social interactions (Larsen-Freeman, 2012). Therefore, even two learners' language development might show entirely different developmental trajectories. Although DST considers the social and cognitive aspects of language development as well, it does not involve the innate linguistic principles for creative language use, but that of human disposition for language learning (De Bot et al., 2007; Griffiths, 2018; Lowie & Verspoor, 2019). De Bot and colleagues (2007) and Larsen-Freeman (2012) also emphasize the importance of the initial state (regarding the first language as well) in language development since a slight change in these variables might result in remarkable differences in the developmental paths. As they posit it:

> 'Regardless of their initial states, systems are constantly changing. They develop through interactions with their environment and through internal self-reorganisation. Because systems are constantly in flow, they will show variation, which makes them sensitive to specific input at a given point in time and some other input at another point in time.' (De Bot et al, 2007:8)

In De Bot and colleagues' (2007) concept, communication is an inter-individual and multimodal action in which complex structures emerge from simple ones from time to time, hence its progress is unpredictable. Multimodality of communication (voice, pitch, rhythm, gestures, facial expressions) directs the meaning-making process, in which meaning is constructed rather than achieved. The change in the system of language development highly depends on those limited and interlinked internal and external resources that might be compensatory. Internal resources like memory, learning style, anxiety, self-confidence, time allocation, level of conceptual knowledge and motivation and external ones like the level of support that can be given by the environment (motivating, language-rich, expansible spatial language reality, and the learning material) all have impact on the whole process. The availability of resources not only has an impact on the learners' language development, but on each other as well. Growth (improvement) in the language system is seen as the change between the levels of present and previous development, considering that the latter one serves as a basis for the actual level (de Bot et al., 2007, 2012; Verspoor & Hong, 2013). Jessner (2008) also claims that the same ongoing changes can be observed in case of third language acquisition. She introduces the notion of the M-factor (multilingualism factor) and defines it as a property that is emerged due to the catalytic effects of the interactions among open language systems. However, it is neither seen as a result, nor a precondition of this dynamism (Jessner, 2008).

In conclusion, the unpredictable individual developmental trajectories caused by the high variability of the initial conditions and the availability of internal and external resources in individuals, call for a careful selection of research methods and data analysis that considers development holistically. Hence, case studies might reveal the actual interaction of specific variables (de Bot et al., 2007).

1.2.3 Usage-Based approaches to second language acquisition

Wulff and Ellis (2018) claim that usage-based approaches hold that the general cognitive mechanisms that underpin any kinds of learning are also employed for language learning. They posit that language involves the learning of conventionalized units, called constructions that can range from single morphemes to more complex phrases. Some of them carry concrete or abstract meanings, while others operate only as function words. The process of language learning is defined as learning associations among and within these constructions. They posit that language learning is a process of exemplar-based statistical analysis from which the language system emerges during usage. The entire process is characterized by dynamism and adaptivity as the consequence of constant interaction of different factors. Learner-related cognitive factors (learnerd attention, automaticity, blocking etc.) and constructions-related factors (saliency, prototypicality, significance of meaning etc.) both seem to be crucial in this process. Since the constructions are stored in multiple forms because of their constant reoccurrence, lexicon and syntax somehow show an overlap.

1.2.4 Content and Language Integrated Learning (CLIL)

CLIL as a usage-based educational programme is in line with the concept under the DST framework. CLIL is theorized as an umbrella term that covers many educational approaches such as immersion, bilingual education, multilingual education, language showers and enriched language programmes (Mehisto et al., 2009). What theorists basically agree on is that "Content and Language Integrated Learning (CLIL) is a dual-focused educational approach in which an additional language is used for learning and teaching of both content *and* language." (Coyle et al., 2012:1). However, this definition is refined by adding a new dimension (procedural choices, i.e., cognitive skills) to the previous two to emphasize the co-existence of thinking abilities triggered by cautiously selected tasks. In their concept, Ball et al. (2015) claim that content and language are also vehicles for subject competences. They draw a parallel between the three dimensions of CLIL and the volumes of a 'mixing desk'. In this analogue all dimensions matter to a certain extent depending on the demands of an activity made salient by the teacher.

1.2.5 CLIL in Europe

CLIL is not a novel educational approach. It was launched in Europe only in the 1990s officially, in response to the European Union's language policy proposals (Kovács, 2006; Coyle et al., 2012; Ball et al., 2015; Goris, 2019). According to the community's proposal, it would be desirable for all European citizens to be able to use two foreign languages in addition to their mother tongue (Kovács, 2006; Goris, 2019). Even though the driving forces for the initiation of CLIL programmes in the European Union might vary from country to country, they are mainly related to the political and educational objectives of the given country. Some countries lay emphasis on the favourable socioeconomic 'by-products' of CLIL, such as future international job opportunities or adaptability to new circumstances, while others emphasize the assumed sociocultural consequences like tolerance and acceptance of different cultures. Another possible reason for initiating a CLIL programme is the learners' language and inter- and intrapersonal development (Mihály, 2009). Regardless of reactive or proactive reasons, countries adopting this programme are in common in their intention to suit the present day demands (Coyle et al., 2012; Attard Montalto et al., 2016; Goris, 2019).

Pérez-Cañado (2012) in a review article on the European CLIL research outcomes concludes that CLIL seems to have supremacy in terms of methodology compared to the conventional language teaching practices of mainstream schools all over Europe. Its popularity is explained not only because of the high level of target language outcomes reported in communicative competence, receptive skills, fluency in speaking and writing, lexical and syntactic complexity, creativity and risk-taking, but those of content knowledge as well. Even though, the related literature emphasizes learners' higher level of L2 proficiency, it rarely considers the diversity of existing programs. Goris and colleagues (2019) in their review article on European CLIL programmes conclude that the cause of inconsistencies in the research outcomes might be linked to the historical-political-structural circumstances in which they were launched, suggesting that studies reporting positive outcomes are from countries where these programmes had been started on a higher initiative.

1.2.6 How does CLIL work?

CLIL is, generally, an additive-type of programme with the aim of enriching the learning environment, but it cannot be separated from the applied techniques of mainstream programmes. Even though it shares many pedagogical and educational practices with other mainstream programmes, they differ fundamentally, since it is content rather than language driven. However, the existing difference among the implementations of CLIL programmes are related to the amount of emphasis laid on content and/or language. In 'hard' programmes the focus is more on content. Conversely, 'soft' CLIL is language-driven (de Boer & Leontjev, 2020). Originally the dual focus in CLIL referred to the parallel development of both content and language, although one or the other is more highlighted on occasion depending on learners' needs. More recently a third dimension was also added. The third dimension of CLIL is procedural skill development as an additional core feature of the programme. This three-way focus of CLIL (developing content and language and the application of adequate processing techniques) is made transparent in Figure 1:

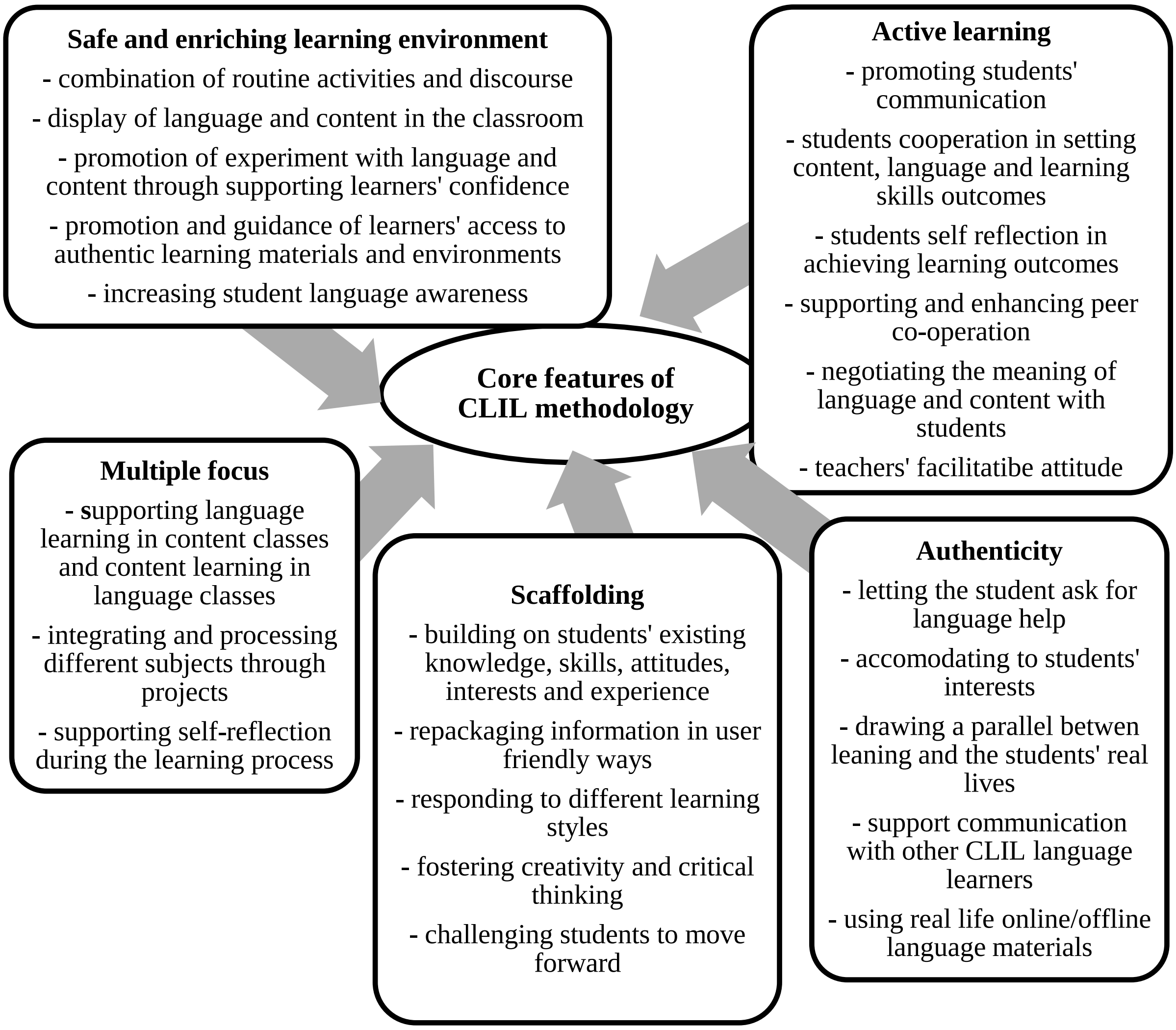


Figure 1: Core features of CLIL methodology (based on Mehisto et al., 2009:29 and Borowiak, 2019.)

The core freatures of CLIL created by Mehisto and colleagues (2009) cover four main principles: cognition, content, community and communication. Cognition refers to higher-order thinking skills like analysing, reasoning, imagining, evaluating, and creating. Content covers not only access to information but linking them to prior knowledge as well. This way, learners create their own content knowledge. The content of lessons in CLIL can be approached by the teachers from many perspectives: they can plan topic-based or cross-curricular lessons that cover the material outlined in the school curriculum and are in line with the learners’ interests as well.

Topics are usually subdivided into smaller sections with both real and not realistic tasks to be completed by the learners. While these tasks are solved, ‘language occurs naturally in the discourse framework associated with the conceptual content, and as a result of the communicative exchanges required by task-based methodology’ (Ball et al., 2015:37). Language of each theme recycles and develops in the same manner from the easiest (e.g., gap filling activities) to the more complex (e.g. oral presentation supported by visuals) ones. During CLIL lessons learners need to be aware of the role the target language has in their local or wider community.

Communication is the main tool to construct knowledge. For this reason, CLIL teachers talk less, and organize their lessons in a way the learners can benefit the most of them (Attard Montalto et al., 2016; Borowiak, 2019). Ball and colleagues (2015) differentiate among three layers of language used in a classroom discourse: subject-specific, general academic and peripheral language. Subject-specific language refers to the special terminology of a certain subject which is necessary for learning about a topic. They are items that have the lowest frequency among those of the three layers. General academic language is not specific to any subject area: it is strongly related to thinking skills. While learners are asked to make comparisons, draw conclusions or create a classification this layer is used. These two layers comprise CALP (Cognitive Academic Language Proficiency) in Cummin’s (2000) approach. Peripheral language is the language of organization or classroom interactions. This is what Cummins (2000) refers to as BICS (Basic Interpersonal Communicative Skills). Unlike CALP, BICS are skills that are used for interacting successfully in social situations. As a more informal language variety, it develops first in a rather quick manner in the first school years. However, neither of the two can be excluded from the school context. As school years go by, the level and amount of abstract knowledge also increases so the focus shifts from one to the other (Cummins, 2000). Making the material of importance clearly visible is crucial in CLIL. More recently a fifth principle has been given to the core principles: competences, referring to the importance of the outcomes of the lessons (Attard Montalto et al., 2016).

Language in CLIL is considered differently than in mainstream programmes. First, it is not seen as a rule-governed, rather a lexically driven system in which syntax subserves meaning. Second, it is immediately used while being embedded in a student-friendly and motivating content (Mehisto et al., 2009; Borowiak, 2019; Goris, 2019). Third, language learning in CLIL is rather incidental than intentional, which means that it is characterized by a more peripheral attentional focus. This is the reason why iteration has a huge role in the language processing, since its evolution moves from focusing on meaning to form (Larsen-Freeman, 2012). The higher the number of occurrences of specific structures and phrases in purpose-based tasks is, the more secure conceptual and language learning takes place (Coyle et al., 2012; Goris, 2019). Constructions of high frequency within the input are more likely to be acquired first by the learners in comparison to those of low frequency, since language knowledge is largely based on statistical learning (Ellis, 2015). As she points it out: 'In other words, what is learned through iteration are not simply meaningful patterns, but the process of shaping them appropriately to fit the present context' (Larsen-Freeman, 2012:204).

Language acquisition should also be supported by the teacher via scaffolding, mediating activities and carefully planned repetition, as well (Lasagabaster, 2008; Wulff & Ellis, 2018; Verspoor & Hong, 2013). Even if the whole system (subsystems of form and meaning) is experienced at the same time by all participants in the language group, learners are expected to process it at different levels, although certain overlaps might take place. Therefore, teaching and learning as such, under the framework of Dynamic-Usage-Based (DUB) approach, cannot be described as a linear process, rather specific conditions provided by the teacher to enhance learners' self-organization (Verspoor & Hong, 2013). Pfenninger and Singleton (2019) posit that CLIL is highly suitable for primary school learners, due to those implicit learning mechanisms that underpin it. Since implicit learning is more alike natural acquisition, learners at lower ages are more attuned to acquire the material instead of learning it in a systematic or analytic way than older learners. For this reason, the distinction between implicit and explicit learning is crucial since the developmental stages that underpin the learning processes can narrow or expand the range of teaching and assessing techniques (Lasagabaster, 2008; Kovács, 2009; Navracsics, 2010; Nikolov, 2011; Saville-Troike & Barto, 2012; Ellis, 2015; Pfenninger & Singleton, 2019).

Ellis (2015), however, draws attention to the fact that not all language input can be taken in. For this reason, naturalistic L2 acquisition cannot be as successful as first language acquisition. He claims that blocking is responsible for directing attention to novel input. Although, the expected association cues are based on prior experience that makes the intake of novel associations difficult. For this reason, explicit teaching methods should also be considered during L2 instruction to direct learners' selective attention to specific aspects of L2. This way intake might be more successful and durable for second language learners who receive limited input even in language-rich environments (Ellis, 2015; Wulff & Ellis, 2018).

1.2.7 Assessment in CLIL

Assessment is another crucial issue in education that can be realized at the level of society, community, school management or the individual. In CLIL it is even more complex due to the different objectives of the programmes. Furthermore, evaluation practices are varied from school to school because of the lack of well-established evaluation criteria, since programmes are bottom-up initiatives and the criteria of assessment are continuously developed and refined during the program (Ball et al., 2015). Regarding the fact that teachers' educational beliefs are manifested in their assessment strategies and techniques (Hercz, 2007), it can be claimed that CLIL teachers' usage-based practices call for the application of different assessment techniques in comparison with those of teachers at mainstream schools (Kovács, 2018). As Ellis (2015) points it out the knowledge gained in either implicit or explicit way should be assessed and tested accordingly. De Boer and Leontjev (2020) claim that the general purpose of assessment is gaining information about learners' skills. This goal 'defines what information is obtained, how it is obtained, how it is *interpreted,* and more importantly, how it is *used*' (de Boer & Leontjev, 2020, pp. 10). For this reason, before introducing a new unit, it is essential for learners to know what the focus is on: language (linguistic form) or content (terminology) or effective communication, and inform the learners about the assessment criteria or the form of assessment (summative or formative, self-assessed, peer-assessed). Learners' improvement is guaranteed, and motivation is also fostered in the long term if self-evaluation is handled as a natural component of the learning process in which the achievement of personalized learning goals is seen more important than long-term curricular goals (Kovács, 2009; Coyle et al., 2012; de Boer & Leontjev, 2020). Hercz (2007) also highlights that the main objective of teachers' assessment is to promote learners' personality, skill, and competence development. The formative aspect of assessment in CLIL is in line with this, since learning in it is process-based and not limited to the improvement of the four basic language skills (reading, listening, writing, and speaking). This is what de Boer and Leontjev (2020) refer to as assessment *for* learning, that is, the main aim of design and practice of assessment is to subserve the individual's development.

Dealing with realistic tasks and creating products like portfolios and projects are more in line with this approach and are also typical and frequent in CLIL programmes. Portfolios and projects are frequently applied devices that are suitable to observe characteristics that cannot be examined in an exact way with traditional quantitative methods (Kovács, 2009). A portfolio is a cautiously compiled collection of works, purposefully created to meet specific pre-agreed criteria, and is suitable to present systematically learners' development as well as to improve their creative ability. Many types of portfolios exist depending on their purpose: presentation portfolio, diagnostic portfolio, cross-curricular portfolio, topic-based portfolio, research portfolio, etc. The integration of portfolios in the learning process has numerous advantages: it might contribute to deep learning, support learner autonomy or groupwork, promote the development of real self-image and motivate (Falus & Kimmel, 2003; Kunschak, 2020). In the process of compilation, the learners have the chance to reflect on their own work and select accordingly (Hercz, 2007). Portfolios can be digitally compiled as well (Kiss-Tóth-Komló, 2008). If they contain works from a certain learning phase, the complexity of learners' development can only be revealed. During their assessment, the teacher is expected to highlight the strengths and the areas to be developed (Falus & Kimmel, 2003; Kunschak, 2020).

Problem-based learning methods that involve performing complex tasks are variants of the project method and are realized in a specific product. Better results achieved by learners who are taught with problem-based methods confirm their effectiveness in comparison with traditional methods. This teaching method is in line with the constructivist pedagogy that highlights the importance of learners' knowledge construction. This method is ideal with small groups of learners of mixed abilities, in which individual developmental pathways are of primary importance (Dancsó, 2007). With the integration of content and language during assessment learners are oriented towards a more holistic and realistic language use (Kunschak, 2020).

Kovács (2018) defines CLIL as the manifestation of constructivist pedagogy. She posits that in CLIL conceptual and emotional development are also in focus. The structure of effective CLIL lessons is reversed. Learners' first produce the language, then practice and finally there is a presentation phase (Kovács, 2009; Ball et al., 2015). Since content organization, lesson structure and classroom arrangement are not comparable to those in the traditional language lessons, the atmosphere is also different. Given that direct correction of mistakes is not typical in CLIL, learners are open to communication. These special features might support maintaining learners' motivation (Navracsics, 2008; Kovács, 2018). Nikolov (2011) highlights also confirms the importance of motivation in the learning process, which she thinks can be maintained with ease if the integration of short and cognitively challenging tasks are instantly followed by individual assessment. It is absolutely in line with the concept that the foundation of a knowledge base does not have priority in CLIL, but life-like use of language does (Coyle et al., 2012).

1.2.8 Why is CLIL suitable for the 21st century?

The 21st century is the age of rapid change and accelerated development. These challenges cover every aspect of human life: social, economic, and environmental. For this reason, education has a crucial role in providing learners with a kind of knowledge that goes beyond curriculum topics. These are the competences that learners can make good use of in their lives and for global goals. The three key competences are as follows: cognitive, interpersonal, and intrapersonal competences. A key competence is defined as a learner's capacity to act in accordance with the upcoming demands to perform well in more complex tasks. The key competences cover skills, dispositions, attitudes, values, and knowledge as well.

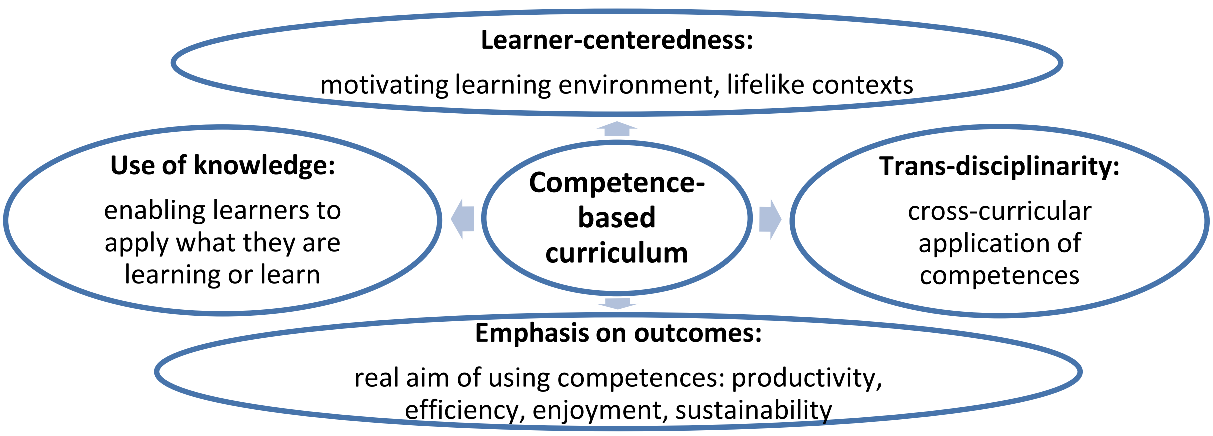


Figure 2: Main principles of a competence-based curriculum (based on Marope et al., 2017)

Cognitive competences refer to cognitive skills like reasoning, critical thinking, creativity, problem-solving, decision-making, handling information, dispositions like curiosity and persistence, values like desire to seek challenges. Interpersonal competences are socio-emotional skills like cooperating with others, maintaining relationships, and influencing others while being empathetic, caring, transactional and trustworthy. Intrapersonal competences are related to self-awareness (strengths and weaknesses), self-regulation (at the level of behaviour and thoughts), and readiness to improve, make effort and be adaptive (Attard Montalto et al., 2016; McGuinness, 2018). To make learners master these competences, a competence-based curriculum is required. Such a curriculum places the learners in the focus of the teaching process. The main principles of the competence-based curriculum are seen in Figure 2.

This paradigm shift in education places the teaching-learning process on a new base. Consequently, in this process teachers and learners' roles should be redefined. Competence-based teaching requires an adaptive teacher whose role is to manage the settings rather than direct them and to transfer knowledge through meaningful practice. Rather than serving ready-made theories, teachers need to focus on questions and problems hence promoting individual and co-operative problem solving (Trilling & Fadel, 2013; McGuinness, 2018). CLIL principles and methodology are in line with the above-mentioned demands: instead of controlling the process, the teacher supports and mediates learners' knowledge construction, therefore they learn the material while it is in use (Kovács, 2009, Coyle et al., 2012; Borowiak, 2019). In this way learners become responsible for their own learning process in which they play an active role.

> 'From this perspective, CLIL not only promotes linguistic competence, it also serves to stimulate cognitive flexibility. Different thinking horizons and pathways which result from CLIL, and the effective constructivist educational practice it promotes, can also have an impact on conceptualization (literally, how we think), enriching the understanding of concepts, and broadening conceptual mapping resources. This enables better of different concepts and helps the learner advance towards a more sophisticated level of learning in general' (Coyle et al., 2012. p. 10.)

1.2.9 Factors influencing success in CLIL

The reasons of the popularity of CLIL programmes are many-folded: it is supported by fundamentally different teaching methods (Coyle et al., 2012), and this results in qualitatively different language knowledge (Kovács, 2018). Attard Montalto (2016), Kovács (2018) and Borowiak (2019) also refer to CLIL programmes as alternative ones, since they are learner-centred and provide their participants those specific thinking skills and learning strategies that they will need in the future. These are some of the reasons why of the high popularity of CLIL programmes in Hungary are at the leading edge on a European scale (Trentinné, 2014).

However, there are factors that might have impact on the success of CLIL programmes and that are not often emphasized. The qualitative difference between CLIL and mainstream programs do not derive simply from the number of years dedicated to L2, since improvement is slower and extended. In the early school years, language development is embedded in a holistic teaching-learning environment therefore no direct (explicit) language teaching occurs. Meaning is in the focus of instruction; hence success highly depends on the clear specifications of long and short-term curriculum objectives that must be in line with the characteristic features of the age groups (Kovács, 2009; Goris et al, 2019).

Learners' socioeconomic status has also been found to crucially impact learning outcomes (Verspoor et al., 2015; Dallinger et al., 2016). So does teachers' L2 levels among which high diversity can be explored Europe-wide (Goris et al., 2019). Time allocation, the language syllabus, L2 exposure and finally learners' literacy and cognitive skills also have an impact on teaching-learning outcomes (Cummins, 2000; Ball et al., 2015; Goris et al, 2019). It is often cited in research that learners with good L1 literacy skills and academic language proficiency have the goods on learning in L2 since underlying academic language skills learnt in one language, can be transferred to the other. This is what Cummins (2000) refers to as 'Developmental Interdependence Hypothesis'. If learners' academic language ability is low and they are unable to meet the language and concept demands, effective learning cannot occur, except their subject knowledge is highly grounded. If learners from a lower socio-economic background participate in CLIL programmes, they usually need more support, since their CALP might not be at the required level in their L1. High target language exposure in out-of-school settings can also contribute to an advanced level of language knowledge (Ball et al., 2015; Attard Montalto et al., 2016). In case of learners with low L1 literacy, Pfenninger and Singleton (2019) suggest the increase of L1 instruction before the initiation of L2 since learners cannot benefit from the interdependency of L1 and L2. However, they also add that there is the problem of thresholds to be defined as the necessary level for L2 initiation, since it can be different for learners. As they finally point it out, the interdependence of literacy skills at least should not be taken for granted (Pfenninger & Singleton, 2019).

Another issue that is closely related to quality assurance is the lack of trained CLIL teachers. Quality language input and teacherly feedback are crucial in CLIL since they primarily support learners' language development, especially in case of younger children. Moreover, content and language teachers' L2 use cannot be restricted to structures that are strictly outlined in traditional language books, even if their order of acquisition is scientifically described, because the language required for expressing a particular content is not always available for learners at the time of teaching. Since meaning is the driving force in CLIL, learners regard language as a meaning-making process (Escobar Urmenta, 2019). Moreover, the language levels required by the teachers to work in CLIL programmes is dissimilar Europe-wide (Mehisto et al., 2009; Van Mensel et al., 2019; Goris et al., 2019; Borowiak, 2019).

1.2.10 Hungarians' foreign language knowledge

Pedagogical culture in the Hungarian public education is often described as conservative, knowledge-based and teacher-centred, although a novel attitude towards second language teaching has been present for some time. Language teachers in Hungary, however, have better opportunities compared to subject teachers in terms of refreshing their methodological practices, since various trainings are held, modern books are accessible as well as good online communities, live worksheets and videos etc. Nevertheless, many of them still work with outdated materials and do not keep up with the needs. Therefore only a very slow change can be experienced (Einhorn, 2015b). Despite the advent of the Communicative language teaching method in the 1980s, many language teachers still rely on the Grammar-Translation method that marked the 1970s. This may be due to the norm oriented curriculum design that narrows teachers' room for maneuver. Teachers often focus on language exam outcomes instead of functional language use, error correction instead of development evaluation (Navracsics & Molnár, 2017). However, this latter one could be a catalyst for effective learning and should be based on differentiation. Csépes's (2019) results on the applied assessment techniques of practising teachers revealed that they are not confident enough in alternative methods (portlolio, self-reflection, peer-assessment, communicative competence and learners with special needs) and they feel the urge to be trained in them.

Eurostat data on the number of known (used) languages in 2016, reported that 57.6% of the Hungarians (aged 25-64) did not speak any foreign languages. Only Albania, Romania, Bosnia Herzegovina and the UK lagged behind Hungary with a higher proportion. Approximately 28% of Hungarians spoke only one foreign language. With this result Hungary was the 25th in the list of 38 countries. The proportion of Hungarians using two foreign languages was 11.1%. Data from Albania, Romania, Bosnia Herzegovina and the UK showed lower proportions. In terms of Hungary, only 2.7% of the people reported knowledge of three or more foreign languages. Regarding younger adults (aged 25-34) the results are slightly better: 40.4% of them reported no foreign language command, 39.7% that of one language, 17.2% that of two languages and 2.7% that of three or more languages. In comparison to other EU countries, with these data Hungary achieves the weakest results. Not surprisingly, researchers found high correlation between the level of education and that of self-reported proficiency among the data. However, when evaluating these results, we need to consider a few circumstances: first, these data were collected from self-reported questionnaires; second, other countries also must cope with foreign language-related problems and third, many leading countries are naturally multilinguals. In general, it can be claimed that many attempts have been made in Europe to improve foreign language teaching: one of them is the initiation of bilingual programmes (Einhorn, 2015a). English is generally the standard L2 for primary school learners in most EU countries. According to Eurostat data from 2018, 44.3% of the Hungarian primary school children learn English, 19.9% learn German and only 0.3% French as their L2.

The report on *Europeans and their Languages* (2012) reveals that the reasons why Europeans learn a second language were also various. Hungarians approached the issue in a pragmatic way: primarily they listed job or study-related reasons for second language learning. Openness to other cultures or to the use of L2 in private lives did not characterize them. On the contrary, the opinion of other participating countries reflected preference of activities related to the establishment of socio-cultural and inter-cultural relationships. Figure 3 reports on these data:

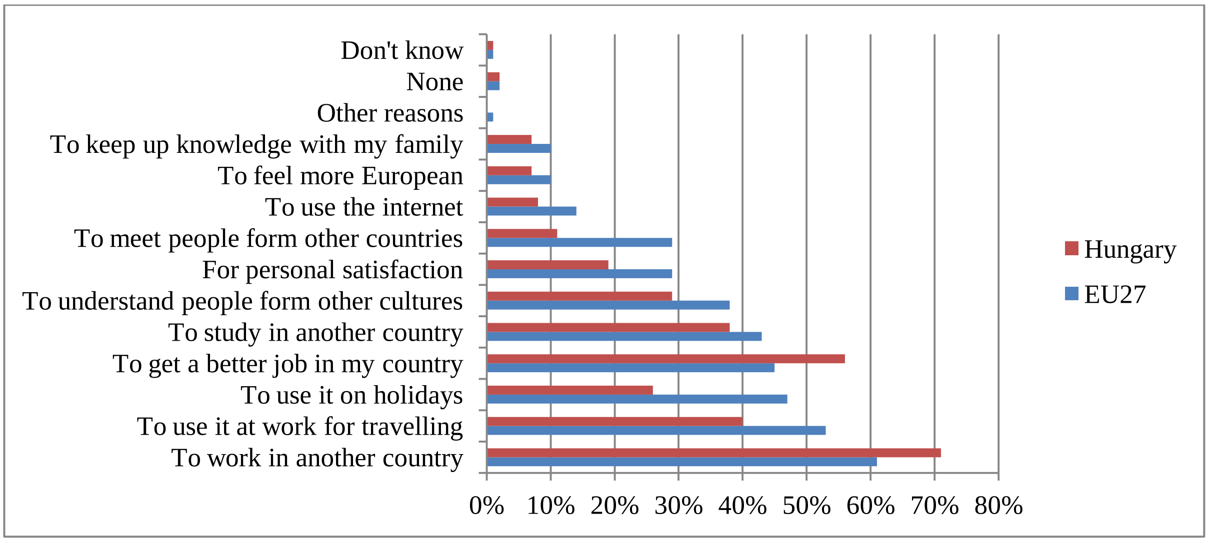


Figure 3: Reasons of L2 learning

(based on Europeans and their Languages 2012:62)

Participants were also asked to list activities that they normally do in their L2s. Figure 4 reports on these data:

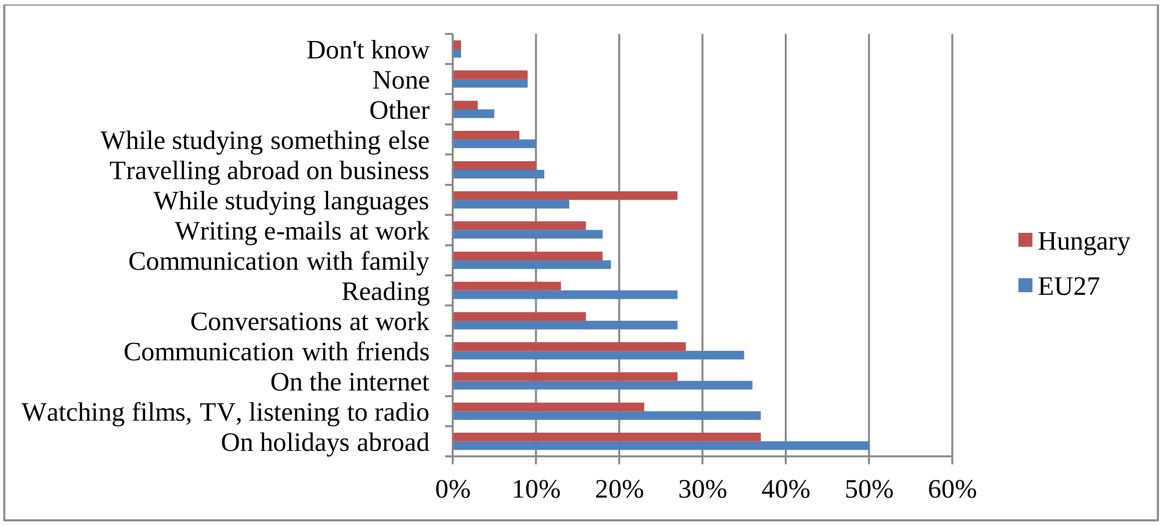


Figure 4: Activities done in the second language (based on *Europeans and their Languages*, 2012:52)

Data show that outstanding results on L2-use are related to free time activities and entertainment for all EU countries. However, Hungarians seem to use their L2 for these activities less frequently as compared to other countries.

1.2.11 Legal background and objectives in second language teaching in Hungary

1.2.11.1 General objectives in foreign language teaching

Recently, the Ministerial Decree 5/2020. (I.31.) has modified the corresponding decree on the National Core Curriculum (NCC). In terms of language learning the new curriculum considers new aspects of language development, such as the apprehension and acceptance of different cultures and the promotion of inclusive behaviour. It also lays the emphasis on the importance of individual multilingualism to which instructed language teaching can strongly contribute. If teachers are ready to rely on students' previous knowledge, language competences and learning strategies and if they are also able to master cross-language teaching methods, they build the foundation of their language awareness thereby preparing them for the acquisition of additional foreign languages. The focal point of language development is functional language use that is the ability to communicate in accordance with one's aims in real situations. It is also essential to explore learners' out-of-school language learning experiences and individual differences on a regular basis to have a deeper understanding of their needs and interests. Integration of traditional and digital channels must be an everyday teaching practice.

The curriculum also specifies the earliest possible (first school year) and mandatory date (fourth school year) for the initiation of the first foreign language. Teaching of a second language can only be introduced after the eighth grade. In the early primary school years language learning should be embedded in playful activities, authentic songs, and materials to maintain learners' motivation. In the upper grades functionality is given even more emphasis with the further expansion of learning spaces. Until the sixth-grade learners are assumed to achieve level A1 according to CEFR (Common European Framework of Reference for Languages). By the end of the eighth grade the output requirement is A2.

Official testing of learners' foreign language progress is carried out annually for sixth and eighth graders. The timing of the procedure is declared in [(Ministerial Decree 27/2020. (VIII.11). In the test papers learners' functional language use is measured, therefore the focus is not on language form but on meaning. For this reason, texts that are used for these tasks are adopted from authentic or near authentic sources. Topics are adjusted to the interests and general knowledge of the age group. However, communication is not assessed during testing. At the end of the sixth grade, learners are required to achieve the minimum level (60%= eighteen scores) to pass the test. Both reading and listening skills are measured in during the one-hour testing session. Tests for the eighth graders cover the same skills. Three tasks are assigned to each skill for which 20-20 scores can be given. The minimum score to pass the exam is 24 scores (60%). The tasks measure whether the learner can understand short and simple spoken or written texts and filter out the necessary information from familiar topics.

1.2.11.2 Main objectives in CLIL

The first [26/1997. (VII.10.)] MKM decree on the issuance of guidelines for bilingual education settled the legal status of schools operating bilingual programs. This one, and the following decrees [(Ministerial Decree 20/2012. (VIII. 31) and 4/2013. (I. 11.)] set out the main objectives for launching the program and the criteria that should be met to maintain it. Recently, the Ministerial Decree 5/2020. (I. 31.) has repealed the previous regulations and new directives and a framework curriculum have been released. They also cover the main competences, cross-curricular links, topics on target language culture, number of subject lessons, communicative and linguistic content that should be considered in the program.

In these documents CLIL is viewed as a holistic teaching approach with the dual focus on parallel development of content and language to achieve balanced bilingualism. In line with the different teaching and learning methods applied in CLIL, qualitatively different requirements are declared in the decrees. In the long run, students learning in CLIL programs should be prepared for being able to process information, learn and work in both languages while their personality and self-confidence develop in a well-balanced way. The daily use of the target language, which is experienced in a close-to-naturalistic language context, makes students capable of thinking in the target language and forming a positive attitude towards people with different cultural backgrounds. The application of various work forms and the integrated development of the four language skills (speaking, writing, reading, and listening) promote the formation and expression of oral or written opinion. The program accomplishes its purpose if students can recognize the positive outcomes of it: the knowledge they gain goes beyond subject-related knowledge and, as a result, novel sources of information are available for further use.

1.2.11.3 Early dual-language educational programmes

Two types of bilingual programs are differentiated in the regulations: programs that require a preparation year (with high number of target language lessons) and those that do not. Since the programs offering a preparatory year are initiated in the secondary education and the focus of the dissertation is on primary CLIL learners, only the program structure with no obligatory preparation year is detailed here.

According to the 4/2013. (I.11.) Ministerial Decree on the issuance of principles of bilingual school education, only those schools are allowed to start a bilingual program that can guarantee it at all levels from the first school year. Target language subjects must be taught in groups instead of classes to ensure quality language environment and outcomes. This rule does not apply for skills subjects. In addition to the target language, at least three school subjects should be taught in the target language. Primary schools decide on school subjects taught in the target language in accordance with the certain rules detailed in the NCC and the subject-specific curriculum framework. According to these rules they must consider the compulsory and freely available number of lessons for the different school years, the priorities of the school and the availability of subject teachers. If History is taught in the target language, subject contents related to the history of the Hungarians are ordered to be taught in the Hungarian language.

According to the curriculum framework for bilingual programs, Civilization should be taught once a week. While designing the subject content, students' age and interests should be considered. During the learning process students become able to draw comparisons among the characteristics, the ethical values, and the culture of the studied countries with those of the Hungarian culture. The interpretation of different social norms and customs like polite behaviour, body language, humour, dialects, and the proper use of vocatives might contribute to a deeper understanding of the culture; therefore it is a more effective communicative strategy. By the end of the 8th grade, students are aware of the most relevant historical events, cultural values, geographical, economic, and political features of the target language countries. They also get an insight into the most important festivals and the typical sports and sport events. The topics provide more and more opportunities for individual research or group presentation with the aim of synthesising the acquired knowledge, form opinion, differentiate opinion from fact, use authentic and digital sources and draw consequences. As a result, students become open to the world and more tolerant in their interpersonal relationships.

1.2.11.4 Entry and output requirements in CLIL

The effective functional use of the target language, as the main objective of CLIL programmes, requires a wider time frame which can only be ensured by an early start and frequent exposure to the language. In the Hungarian educational system, instead of entrance exams, output requirements are regulated. Although, CLIL programmes in Hungary aim at developing content and language in parallel, surprisingly, only L2 skills are measured. Moreover, communication in CLIL has a highlighted role in the programme, it is not assessed at all.

The timing of the procedure is declared in [(Ministerial Decree 27/2020. (VIII.11)]. The levels of language skills are specified in the curriculum framework for bilingual programs which is in line with the CEFR (Common European Framework of Reference for Languages) levels. The quality of target language development can only be guaranteed if students are offered four language lessons a week until the end of third grade (144 lessons per school year) and five language lessons for the remaining school years (180 lessons per school year). The Ministerial Decree 4/2013. (I.11.) regulates the language levels to be achieved by the end of certain school years, that is seen in Table 1

	School year			
	2nd grade	4th grade	6th grade	8th grade
CEFR level	cannot be specified	A1 (Not tested)	A2	B1

Table 1: Language levels to be achieved by the end of 4th, 6th, and 8th grade

The purpose of target language examination is to assess the ability of functional language use, (communicative competencies). Therefore, students are required to solve tasks like real language use. Since the focus is rather on meaning than linguistic form, the texts used in the task sequences are authentic, close to authentic, or adapted from an authentic source (e.g., stories, advertisements, emails, blog posts, newspaper articles, interviews, or dialogues).

The topics of the tasks are adjusted to the knowledge and interests of the given age group and correspond to the recommendations of the National Core Curriculum and the Framework Curriculum. At B1 the compilation of tasks covers three basic language skills: listening comprehension, reading comprehension and writing. The listening comprehension task might give an insight of the student's level of understanding people having an everyday conversation in the standard and clearly articulated variety of the target language. The two reading tasks focus on the student's ability of understanding authentic texts written in everyday language on filtering out information from more complex texts. One of the writing tasks assesses whether the student can produce a simple, coherent text in each topic. In the other task, the student is required to write a formal letter of complaint or information or an informal letter about an event or individual experience. Table 2 shows the structure and the minimum requirements of the target language exam:

Language skill	Allocated time (minutes)	Number of tasks	Maximum score	Minimum score to be achieved
Listening comprehension	30	2	20	36 (60%)
Reading comprehension	30	2	20	
Writing	30	2	20	

Table 2: Structure and minimum requirements of the target language exam in Hungary, from:https://www.oktatas.hu/kozneveles/meresek/celnyelvi_meres/eredmenyek

Instead of individual performance, mean scores for each school grade is calculated and observed. These written test results have an impact on the future operation of the programs: if less than 60% of students at a certain grade can achieve the minimum level within a period of three years, the program will be discontinued. Even though the series of tasks are levelled by professionals and experts, they are not piloted or standardized. Furthermore, the compilations are corrected locally, by the teachers at the school; hence, no direct conclusions can be drawn from the results.

Kovács (2018) reveals the contradiction between the main objective of CLIL (improvement communicative competence) and the way it is formally assessed. Given that, there is no formal assessment device of communicative competence provided, it is up to the teachers' conscience whether students achieve the final level goal of B1 in their oral communication. Level of key competences (range, accuracy, fluency, interaction, and coherence) detailed in CEFR (Common European Fraework of Reference for Languages) serve as guidelines for teachers as final requirements for learners by the end of the 8^{th} grade.

1.2.11.5 Target language exam results

Data presented in Table 3 confirm the growing demand for the institutions at a primary and secondary level organizing a bilingual program. English, German, and Chinese are the taught languages in the bilingual programs in primary education. At the primary level, the rate of eighth graders learning in a Hungarian-English bilingual program is the highest compared two the remaining languages. The number of institutions and learners involved in the target language measurement are distributed in Table 3:

	2014	2015	2016	2017	2018	2019
Number of institutions at primary and secondary level (English, German, Chinese)	79	86	91	104	103	105
Number of students at primary and secondary level (learning English, German, Chinese)	3138	3783	4266	4807	5149	5812
Number of 8th graders (learning English, German, Chinese)	1190	1496	1727	1961	2168	2482
Number of 8th graders of English	886	1124	1322	1494	1763	2030

Table 3: Number of schools and learners involved in the target language exam from 2014 to 2019, from: https://www.oktatas.hu/pub_bin/dload/kozoktatas/meresek/celnyelvi_meresek/Celnyelvi2019_Eredmenyekosszesitese.pdf

Eighth graders in Hungarian-English bilingual schools achieved an average of 49.7 to 52 scores between 2014 and 2019 and of whom more than 90% met the minimum criterion of 60% each year. Figure 5 reports on the average scores achieved by the eighth graders from 2014 to 2019, the proportion of students who met the minimum criterion and the results gained in the different task types:

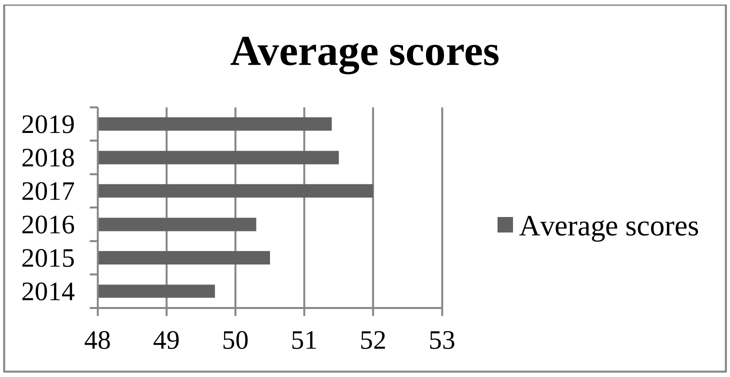


Figure 5: Average scores achieved from 2014 to 2019, from: https://www.oktatas.hu/pub_bin/dload/kozoktatas/meresek/celnyelvi_meresek/Celnyelvi2019_Eredmenyekosszesitese.pdf)

Results show that CLIL learners achieved the highest scores in the target language exam in 2017. Since then, a slight decrease can be observed, however the average rate is still above 85.5 percent. Figure 6 reports on the proportion of students who met the minimum requirements:

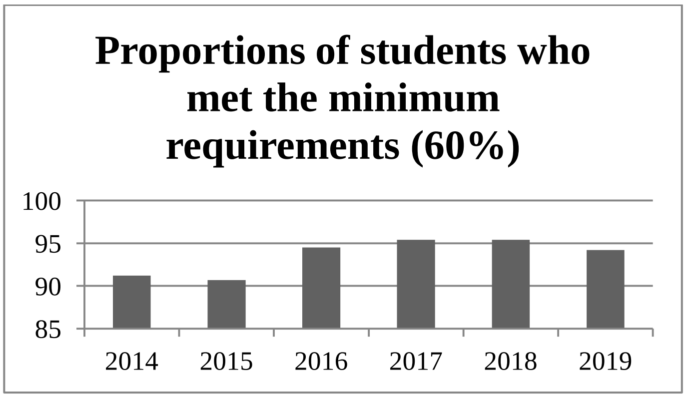


Figure 6: Proportion of eighth graders who met the minimum requirements, from: https://www.oktatas.hu/pub_bin/dload/kozoktatas/meresek/celnyelvi_meresek/Celnyelvi2019_Eredmenyekosszesitese.pdf)

The number of CLIL learners achieving the minimum level in the target language exam increased until 2017 and it did not change remarkably for a year. Since then, a slight decrease could have been detected. Figure 7 reports on the average results of different English language skills achieved each year:

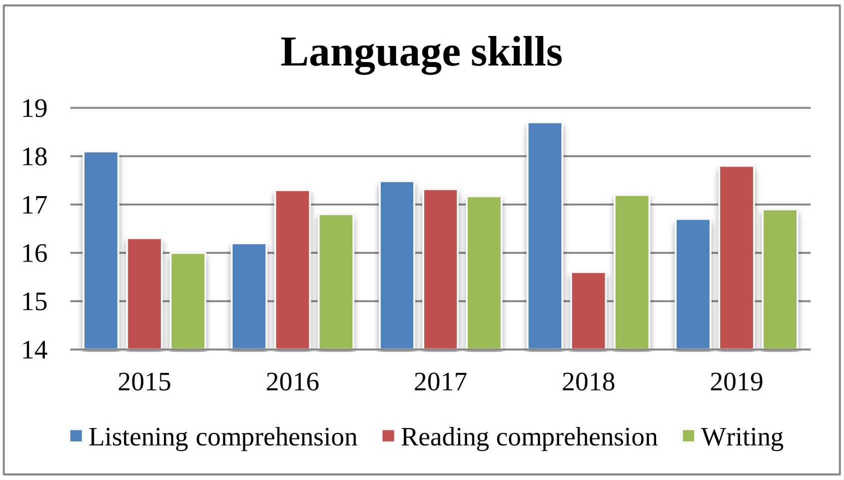


Figure 7: Eighth grader (CLIL) learners' language skills, fromhttps://www.oktatas.hu/pub_bin/dload/kozoktatas/meresek/celnyelvi_meresek/Celnyelvi2019_Eredmenyekosszesitese.pdf)

During the years, surprisingly the results on the productive (writing) component of the exam have shown the slightest difference in comparison with the other language skills.

1.3 Executive functions and their components

Language processing and development cannot be investigated without considering the underlying cognitive mechanisms and the related constructs. Therefore, in the following sections previous and current conceptual frameworks on the executive functions (hereinafter EFs) and the bilingual mental lexicon are detailed.

The development of executive functions has a huge impact on success related to different areas of life such as career goals or the quality of human living and relationships (Diamond, 2016). Even though they are considered as a relevant psychological construct, there is neither consensus on their definition, nor on the exact set of operations covered by them (Győri, 2008). According to Stuss and Alexander

(2000) the reason for this impurity is the inconsistent and interchangeable use of a psychological concept (executive functions) and an anatomical construct (frontal lobe functions). In general, the umbrella term covers those modality-independent, top-down functions that modulate and coordinate mental systems that are responsible for human cognition and adaptive behaviour in a fast-changing context. These functions are attention-regulation skills that are often required in certain context-dependent situations and contribute to behavioural patterns like goal-orientedness, self-reflection, planned, focussed or adaptive behaviour, sustainment of attention, resistance of distraction and understanding the consequences of attitudes (Zelazo et al., 2016). More specifically, these distinct processes (control functions) constitute a central supervisory system which is activated at any time for any types of tasks regardless of their complexity (Stuss-Alexander, 2000).

1.3.1 Previous and current models on executive functions

In the following sections previous and current models on executive functions and its core functions are detailed.

The concept of executive functions dates back to neuropsychological research based on patients' behavioural changes as a result of the damage of their prefrontal cortex. Basically, this brain area was thought to be primarily responsible for the operation of necessary functions of successful life management. Since lesions in this brain area caused various symptoms in patients' behaviour and cognitive functioning such as the lack of the ability of strategic planning, implementation, attention, giving appropriate feedback or thinking flexibly, early approaches regarded executive functioning as a unitary concept and assumed the exclusive involvement of the prefrontal cortex (Győri, 2008).

Debate on the nature of executive functions is related to its unitary or diverse nature. Recent EF approaches support the idea of a psychological construct with many separate, but interrelated processes. However, there is still no consensus on the number of its factors if age-related development is considered. As a result of their factor-analysis, Miyake and colleagues (2000) define three core EFs: working memory, inhibitory control and shifting. They also claim that they are both a unitary and a divisible construct that are interrelated and separate at the same time. However, their method does not cover different age groups, only adult participants (Miyake et al., 2000;

Daucourt et al., 2018). As opposed to Miyake and colleagues (2000) other researchers propose a single-factor (Robinson et al., 2012), two-factor (Messer et al., 2018) or multi-factor structure of EFs (Diamond, 2013, 2016) in case of children aged 6-12. These authors explain the differences among the age ranges with age-related developmental changes. Some of these models exclude inhibition or shifting, while others merge the two into one. Messer and colleagues (2018) propose a two-factor model including a separate (verbal and non-verbal) inhibition factor and a general EF factor (working memory, shifting, fluency, and planning) in case of primary school children. Although their proposal is in line with studies identifying at least two factors in case of primary school children, they do not group children by their ages. Xu and colleagues' (2013), however, define different age ranges in their study and their results vary accordingly. For children between the age ranges of 7- 9 and 10-12 they found a single EF factor. In case of the third age group (children aged 13-15) a three-factor model of EFs (working memory, inhibition and shifting) was confirmed providing evidence for the assumed age-related developmental changes. Diamond (2013) completes the three core functions with higher order functions of reasoning, problem-solving and planning.

More recently, EFs have been considered more separable over time because of natural development, although the exact age ranges at which their factor structure changes from unitary to multifactorial are only assumed (Miyake-Friedman, 2012). The reasons why the different investigations on the structure of the EFs result in dissimilar outcomes might be that, various tasks are applied or the used tasks assess a function which might also be relevant in another function.

1.3.2 The core executive functions

1.3.2.1 Working memory (updating)

Updating is strongly related to working memory function and refers to the ability of monitoring and adapting to the changing context, holding, and manipulating information, linking information with thoughts that have been activated in the long-term memory to perform complex tasks. It is highly linked to academic success: without updating written and spoken language cannot be processed and communication as such would not exist in the L1 or L2. Cognitive processes like recognizing logical connections and drawing conclusions that lead to making decisions are also supported by the working memory. Making connections or associations among seemingly unrelated ideas, which is the basic feature of creative thinking, also needs the involvement of the working memory (Miyake et al, 2000; Diamond, 2013; Diamond, 2016). Working memory capacity covers processing and temporary storage that both rely on the same attentional resources, so when attention is allocated for one of them, it is not available for the other. If the cognitive load increases the executive control of attention decreases, resulting in reduced inhibitory control over intruding stimuli. Working memory is highly engaged in any learning processes that includes manipulation with information. Strategy use in learning might also result in less workload, therefore a more effortful intake of information. Working memory can be enhanced in the teaching-learning process through the application of mnemonic techniques. Keywords, acronyms, pegwords, phrases, chants, songs, short poems, riddles, and visuals also can help students in chunking, retaining and manipulating the necessary information. Since rehearsing is indispensable in the process, sufficient time must be allocated for it. Working memory performance improves as time goes by, since rehearsal becomes automated (Meltzer, 2014).

Research on working memory aims to reveal whether it is a unitary or a multi-component construct. The evolution of the working memory started in the 1970s when it has been distinguished from the short-term memory (STM) referring to their different roles. While STM was seen as a limited capacity store with earlier and faster development compared to the working memory, this latter one was assumed to be responsible for the temporary maintenance and manipulation of information. The decomposition of working memory into different components began with Baddeley and Hitch’s (1974) hierarchical model. However, the main difference among the existing models is the number of components they assume to be included. In Baddeley and Hitch’s (1974) model, they differentiated two constructs: the visuo-spatial sketchpad and the phonological loop. The central executive was responsible for not only their management but the coordination between the components of the two stores. The phonological loop and the visuospatial sketchpad were assumed to support the maintenance of verbal and visual information and serve as slave systems, with the supervision function of the central executive. These subsystems not only store information but process them as well. This model assumes that phonological information is managed in the phonological loop (verbal working memory). However, it only has a limited capacity, but a remarkable role in the development of long-term phonological representations not only in the L1 but in the L2, also. The phonological loop can be further subdivided into two components: the phonological store and the articulatory rehearsal. The phonological store retains the information temporarily, but the articulatory rehearsal makes it accessible with continuous updating. The capacity of the phonological loop develops between the ages of 7 and 10. Although remarkable individual differences might be detected in its capacity, it is seen as a good predictor of people's language acquisition abilities: a greater early-age capacity results in faster first and second language learning that can be detected in the size of the lexicon. This is in line with previous research highlighting a strong relationship between the long-term memory and working memory: Working memory relies highly on the long-term memory. Children achieving higher results in verbal working memory tests are likely to generate longer statements, and fewer errors in tests measuring syntactic ability. However, it is still unclear whether phonological loop is what supports vocabulary growth, or the increased vocabulary improves the extent of it (Racsmány, 2004; Racsmány, 2007; Turi et al., 2014). In 2000 Baddeley added a new temporary store, the episodic buffer to the model. It functions as a multidimensional store to integrate information from the phonological loop, visuo-spatial sketchpad, the long-term memory, and the central executive. He also claims that the components of the model operate simultaneously but separately, that is, each three remain intact in case of damage to the fourth one. The temporarily activated information is linked to the long-term memory via the domain specific stores. Domain-general executive control processes are assumed to coordinate within the stores and to retain the fading links active (Fenesi et al., 2015).

Originally the role of the central executive was to integrate and link information of different modalities to which it has a direct access through conscious awareness. Later its core functions were supplemented with the role of coordinating sub-systems, inhibiting over-learned behavioural patterns, switching between tasks or retrieval strategies, selectively attending to one stimulus while ignoring (inhibiting) others (Diamond, 2013; Janacsek et al. 2009). The selection of lower-level patterns of thought and behaviour are decentralized and directed by the functional sub-systems, while that of less routine-like operations require the involvement of the central executive. In other words: without its proper functioning, monitoring, flexibility, and effective planning of actions would not exist in human behaviour. Since Baddeley did not emphasize the importance of attention in the model, he did not provide information about the exact implementation of the control process as well (Baddeley, 2000; Racsmány, 2004; Racsmány, 2007).

1.3.2.1.1 Alternative models

Although Baddeley’s model is still the most often cited and applied in scientific literature, some alternative models emphasized more the role of the attentional focus as a unitary control component in working memory functions. Engle and their colleagues (1999) were the first who described working memory as a unitary construct being involved in the temporary maintenance, operation, and retrieval of modality-independent information. In this non-hierarchical model working memory is made up of two components: the short-term memory (i.e., temporarily activated visuo-spatial, phonological, or kinetic long-term memory traces) and the controlled-attention function of the working memory (or executive system; executive attention) that is responsible for focussing on certain traces while excluding (inhibiting) others. They claim that working memory is highly linked to attention and control ability, hence greater attentional control results in better working memory functioning. They claim that differences between groups of high and low working memory capacity individuals can only be detected when attention needs to be controlled during a task, if this control is not needed, the difference between them disappears. With other words, if the task can be fulfilled with automatic processing, then individual working memory capacity differences might not be revealed (Engle et al., 1999; Tuholski et al., 2001; Racsmány, 2007; Fenesi et al., 2015).

Another alternative model of working memory is Cowan’s Embedded-Processes Model. As opposed to Engle, he denies the existence of an entire system; conversely, he defines working memory as different memory phenomenon that is made up of three components arranged hierarchically: inactive long term memory traces that can be retrieved by stimuli; active long term memory traces that are intentionally attended or unattended. This model differentiates two levels: at the first level there are several long-term memory representations from which many can be activated at a high level.

However, at the second (embedded) level, which is the level of focused attention, only a few representations are salient. Since in this model there are no underlying slave-systems for processing information of different modalities, hence, the attentional control has a crucial role in placing the stimuli into the focus to make the information retrievable from the long-term memory. With other words, working memory is assumed to be directed by focused attention (Racsmány, 2004; Racsmány, 2007). Although Engle and colleagues (1999) and Cowan (1999) support the idea of controlled attention, they do not explain its functioning.

1.3.2.2 Inhibitory control

Another core function is called inhibition or inhibitory control, which is considered more heterogeneous than working memory. Since various terms are used in different research domains, its conceptualization has constantly been under scrutiny. The most prominent models describing it are one-factor or multi-factor models. While one-factor models assume a single inhibitory factor, multi-factor models not only suggest the fractionation of diverse inhibitory processes (automatic, behavioural, cognitive, effortful, pre-potent, proactive inhibition and inhibition of return) but that of functions as well (resistance to interference or suppression of prepotent response). Two-factor models focus on the general/specific, task-relevant/irrelevant involvement of attention and inhibition in cognitive processes.

Friedman and Miyake (2004) propose a three-factor model of inhibitory control. They differentiate three main factors: suppression of prepotent responses, resistance to interference from distraction and resistance to proactive interference. Even though they found strong correlation between suppression of prepotent responses and resistance of distractor interference, they regarded them as distinct factors. They assume that their strong relation might be underlied by the ability to maintain goal-orientedness to avoid making inappropriate prepotent responses or to filter out task-irrelevant information. Gärtner and Strobel (2021) claim that, the reason why suppression of prepotent responses and resistance to distractor interference are often assumed to be independent from one another is related to the stage of processing. Resistance to distractor interference is considered to occur at an initial (perceptual) stage. Conversely, prepotent response inhibition occurs at a later stage when the focus is on the suppression of automatic and behavioural responses. This way, they seem to be organized hierarchically.

Contrary to Friedman and Miyake's model (2004), Diamond's (2013) proposes a two-factor model of inhibition. She distinguishes inhibitory control of attention (interference control) and that of action (self-control). If inhibitory control of attention is involuntarily detected (when a salient impulse comes) it is called exogenous, bottom-up, automatic, and stimulus-driven. Contrarily, if the attention is directed voluntarily on a particular stimulus, it is endogenous, top-down, active, and goal-driven. This is what we call as selective (or focused) attention. Another distinguishable aspect of inhibitory control of attention is cognitive inhibition and it is in action when previous experience or memories or unwanted thoughts try to gain space in the mind (Diamond, 2013). Inhibition of action covers motor responses and has many subtypes that are related to self-control.

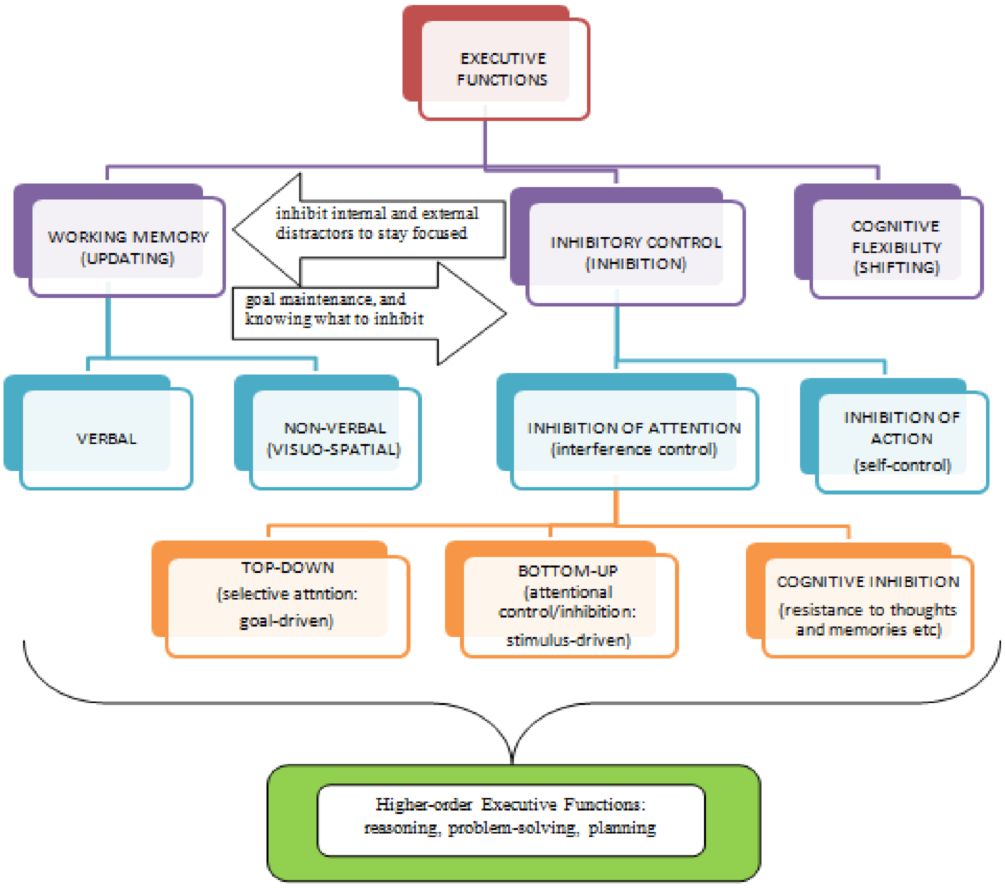


Figure 8: Executive functions (based on Diamond, 2013, 2016)

This makes it possible for humans to override the overused behavioural patterns, resist acting impulsively, keep social norms, be goal-oriented and complete a task despite the distractions or thoughts of giving it up or delay gratification with the certainty of a greater reward later. Impairments in the inhibitory control have general learning, behavioural and social consequences for children, and cause difficulties in memory retrieval and resisting distractions and a slowdown in processing (Howard, et al., 2014; Diamond, 2016).

Howard and colleagues' (2014) two-factor model on inhibitory control based on a factor-analytic method is not in line with Friedman and Miyake's (2004) findings. They proposed that tasks measuring 'prepotent response inhibition' and 'resistance to distractor interference' are performed by the same underlying factor. In line with previous findings (see in Howard et al, 2014), they suggest the automatic-effortful model of inhibition. In this model they claim that effortful mental attention plays a role in hyper-activation of task-relevant schemes parallel with the concurrent (effortful) inhibition of task-irrelevant ones. Automatic inhibition (as a by-product of effortful attention) deactivates task-incompatible mental operations that are not in the centre of effortful attention. Gärtner and Strobel (2021) also replicated Friedman and Miyake's (2004) latent variable analysis. Their results showed only weak correlation among the applied tasks; so they could not confirm Friedman and Miyake's (2004) model, either. In addition, they conclude that tasks used for measuring inhibition should be revised; since many of them require multiple cognitive processes hence they are not suitable to measure inhibition.

Working memory and inhibitory control are highly related constructs. This interdependence has led to the debate on the question whether they are inseparable from each other or they are parts of the same construct (Diamond, 2013; Diamond, 2016; Engle, 2018).

1.3.2.3 Cognitive (mental) flexibility

Another core executive function is shifting (often named as cognitive or mental flexibility). In a broader sense cognitive flexibility (mental flexibility) involves thinking about something in different ways, from different perspectives. It covers the abilities of changing perspectives spatially or interpersonally, shifting approaches or mindsets, adapting to unfamiliar situations, adjusting to the changed demands, and thinking creatively. It is also considered to be crucial in switching between tasks and thinking divergently. Lack of cognitive flexibility (shifting ability) might cause behavioural, learning, and attentional problems. Perseverations are often regarded as signs of deficit in shifting. However, developmental changes during the whole life have impact on cognitive flexibility (Miyake et al, 2000; Meltzer, 2014; Diamond, 2016; Buttelmann & Karbach, 2017). It emerges later in life in comparison to the development of the working memory and the inhibitory control. Adults' faster and more accurate cognitive flexibility is contributed to their improved working memory capacity, interference control and perceptual speed in comparison to children's (Dajani & Uddin, 2017).

Cognitive flexibility highly depends on cognitive processes like salience detection, attention, working memory functioning, switching and inhibition. Salience detection is the most relevant initial step that establishes attention allocation. Both goal-directed (top-down) and stimulus-driven (bottom-up) attentional processes might be involved in cognitive flexibility. Once a change in the current situation is considered, inhibitory control is engaged to stop and override automatic behavioural patterns. At this point new strategies are set so as to promote the necessary shift. Working memory is responsible for monitoring the process (Janka, 2017; Dajani & Uddin, 2017). The processed involved in mental flexibility can be seen in Figure 9:

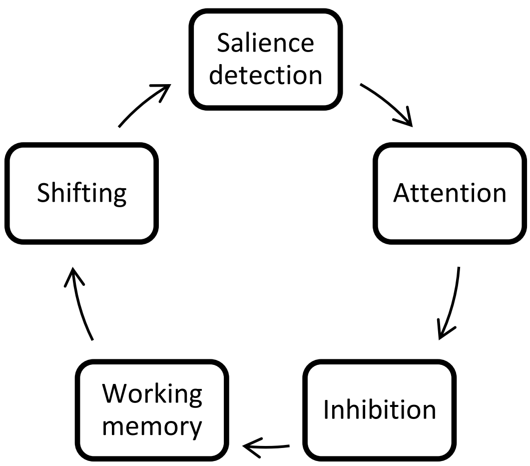


Figure 9: Processes involved in mental flexibility (based on Janka, 2017, p. 1775)

Janka (2017) adds that self-control has a huge impact on shifting behaviour in line with the environmental demands. It has two components: a trait that is an inborn personal characteristic and a state that can be changed. Working in a pleasant atmosphere and experiencing positive emotions are favourable conditions for thinking, focusing attention, working memory functioning, being open to novelties, problem solving or creative abilities. Openness, in addition to being closely linked to creativity, is an essential condition for mental flexibility.

Cognitive flexibility has been often described by two sub-categories (set shifting and task switching) that are also the main aspects by which cognitive flexibility is investigated (Diamond, 2013). During task switching, participants need to switch between diverse instructions/dimensions within the same task (like in the Dimensional Change Card Sort Task). This is quite a challenging task for children aged three or under since they cannot override the first instruction, and they get stuck to it even if they can verbalize the second instruction. In case of adults, reaction times increase when shifting is required. However, set-shifting is related to shifting attention between features (colour or shape) of the stimuli to complete the same instruction. In this test criterion, attention is allocated to switching between the features of the stimuli and not to the instructions. Set-shifting is often described as the less complex process between the two. However, they both cause the slowdown of response times and decrease of accuracy. This slowdown (switch cost) is the result of the less effective functioning of the inhibitory control (Diamond, 2013; Dajani & Uddin, 2017).

As children develop, their abilities and willingness to adopt and shift between unusual approaches also change (Miyake et al, 2000; Meltzer, 2014; Diamond, 2016; Buttelmann & Karbach, 2017). If the context in which language learning is embedded rich in stimuli, students continuously need to monitor and adapt their behaviour; therefore their abilities of changing perspectives and solving various problems develop. Promoting learners' flexible thinking in the learning process this way is a conscious decision made by the teachers. Tasks requiring multiple view-points, collaboration or critical thinking are all contribute to its development. Solving puzzles, forming jokes, putting words into word categories, guessing word meanings, analyzing or interpreting L2 discourses in unusual ways, assuming meaning in specific sayings, attributing attitudes beyond actions, drawing conclusions and writing stories are well-known methods in language teaching (Meltzer, 2014). However, the assumption whether cognitive flexibility can be promoted by context-related cues is not in line with Diamond's (2013, 2016) concept in which cognitive flexibility is characterized by generalizability (that is transferability) that has not been evidenced yet. More recently researchers have suggested the consideration of context-specificity that is the triggering effect of particular circumstances on task performance. Consequently, if response to certain stimuli can be conditioned by the context, then bottom-up processes come into focus (Braem & Egner, 2018).

Taken together, developmental changes and favourable environmental factors can attribute to an increase in cognitive flexibility. For this reason, many different types of intervention trainings have been implemented, like curriculum-based school programmes (Diamond, 2012; Buttelmann & Karbach, 2017).

1.3.2.4 Selective visual attention

Although selective attention has impact on the proper functioning of the EFs, most of the models do not provide detailed explanation on its functioning or lay the emphasis on its role in multiple ways.

Whether top-down (endogenous) or bottom-up (exogenous) attention is engaged, is highly dependent on the nature of the stimulus. If it differs from the distractors only by a single feature like colour or shape (feature-search), the cue is salient, thus attention is driven exogenously. However, if it differs from the distractors in at least two dimensions like colour and shape (conjunction-search), respondents' reaction times increase indicating the involvement of endogenous processes that are linked to executive functions (Friesen et al., 2015).

Working memory and selective, focused attention strongly rely on one another. Working memory trainings seem to support developmental improvements in selective attention (Tuholski et al., 2001; Diamond, 2013). In case of the conjunction-search task, researchers found differences between individuals with high and low working memory capacity. Since increased controlled attention was necessary to complete the task, participants with high working memory capacity processed it more efficiently, thus with lower processing speed.

1.3.3 Factors influencing EF development

1.3.3.1 Cognitive development

Brain development is hierarchical which means that cortical areas that are responsible for the proper functioning of sensory and motor processes develop first. Consequently, executive functions as domain-general cognitive functions are linked to the maturation of the frontal and pre-frontal cortex as compared to other brain regions. Therefore, this dynamic malleability follows a hectic pattern of regression and improvement and actually covers a lifelong developmental trajectory. In this process which crucial and rapid developmental phases can be detected: one of them is that of preschool years and the other is the transition phase to adolescence. The improvement in them is attributed to the reorganization of the neural circuitry in the frontal lobes. As a result, the developing language abilities and social behavioural patterns become more complex in these two age groups. This is the time when attention, impulse control, self-regulation and working memory development also increase. As children enter school their thinking abilities are highly challenged due to the academic requirements posed by educators (Otero et al., 2013). Even though executive functions are involved in school achievement, they are only challenged at the beginning of the learning process, until the behavioural pattern becomes automatic due to repeated practice. After gaining expertise in anything, the involvement of the prefrontal cortex is not needed so newly gained knowledge is passed on phylogenetically older brain regions to allow the process of new information (Diamond, 2013, Zelazo et al., 2016).

1.3.3.2 Living conditions, parental attitude, supportive environment

However, beyond these two most sensitive periods the state of executive functions does not remain permanent: they show continuous development until the early mid-adulthood when their decline starts (Zelazo et al., 2016; Diamond, 2016). On this developmental trajectory many influencing factors might leave their traces such as poor living conditions, stress, depression or sleep deprivation might cause detectable changes in the EF and the prefrontal cortex (Kroll-Bialystok, 2013; Zelazo et al., 2016). Therefore, early intervention in the EF development of children at risk is evidenced to influence their cognitive development in a positive manner. Parents' supportive attitude that is encouraging autonomy and problem solving, ensuring rich oral language environment and avoiding total control over children's behaviour can lead to better EFs. Conversely, unfavourable social conditions (domestic violence, parents' divorce, to be a member of a minority group or under stress) or economic status (poverty, frequent move) can be blamed for poorer EFs and mental or physical problems. Even relatively small drawbacks caused by direct or indirect environmental factors might cause increased differences among children in each passing school year (Diamond, 2016).

1.3.3.3 Language context

Recognizing the limitations of Miyake's (2000) unity and diversity model Green and Abutalebi (2007) suggested the consideration of a new aspect: the determining role of different interactional (language) contexts with variant demands on the adaptability of the cognitive control. They list three everyday interactional contexts of language use that reflect the different demands on the cognitive control:

a) In a single-language context language use is context-dependent so all languages are used separately in different domains, thus there is no need for frequent code switching.
b) In a dense code-switching context people borrow sequences from both their languages even in a single utterance.
c) In a dual language context both languages are used but with different speakers. The use of code-switching in a simple utterance is less typical but might occur in the discourse

According to the strong variant of the adaptive control hypothesis the reason for adaptation of cognitive processes to the ongoing contextual demands is to avoid possible interactional costs and it is not triggered by the reflections made by the context for the inadequate respond as it was previously assumed. In accordance with this hypothesis Green and Abutalebi (2007) refined Miyake and colleagues' (2000) model on executive functions and distinguished eight control processes: goal maintenance, interference control, salient cue detection, selective response inhibition, task disengagement, task engagement and opportunistic planning. In the dual-language context the control processes are challenged the most (all but one process, the opportunistic planning, is highly involved). In order to reduce the interactional costs the speaker needs to constantly monitor the context: to sustain the attention to the language spoken, to suppress interference and to be ready for an immediate shift to the other language (Green & Abutalebi, 2007).

1.3.4 Testing executive functions - measurement impurity

At this point it can be clearly seen that testability of the executive functions at neurological or behavioural level is not less than challenging for many reasons. First, the completion of many EF tasks requires the involvement of multi-factorial executive processes (Stuss and Alexander, 2000; Racsmány, 2004; Diamond, 2013; Paap et al, 2015; Messer et al., 2018). Secondly, not just executive but non-executive cognitive processes might also be touched upon like engagement and motivation that are also necessary in all performance-based EF tasks. Thirdly, task familiarity (retesting) as well influences the test outcomes, while reducing its effectiveness (Miyake et al., 2000).

Fourthly, as Stuss and Alexander (2000) point it out that the participants' unstable mental state might also be mapped in the test outcomes. Furthermore, other hidden factors like the problem of cause and effect, selection or extent might also have influence on the test outcomes. Selection is a two-fold dilemma: firstly, it refers to individuals' personality features and suitability that drive them to do a certain activity, secondly due their choice they are the ones who are selected to take part in experiments. The problem of extent reveals that results of these experiments as evidence for modification are closely related to and limited by the training itself. Although there is considerable evidence that EF skills can be trained, and that the benefits of this training extend to tasks that are similar to the ones used during training, there is less evidence that trained skills transfer to behaviour on less similar non-trained tasks (Zelazo, et al., 2016). Table 4 summarizes those tests that are often applied to measure executive functions:

EF components	Testing method
Planning	Tower of Hanoi, The Tower of London Test
Inhibition	Stop Signal task; Go-noGo task; Stroop test; Eriksen paradigm; Trail Making test
Vigilance	TOVA (Test of Variables of Attention); Integrated Visual and Auditory CPT
Working memory	Backward Digit Span; Listening Span task; Visual Patterns Test; Corsi Blocks Task; Rey-complex task; N-back Task
Cognitive Flexibility	Wisconsin Card Sorting Test; Rapid Automatic Switching; Verbal Fluency Tests, Design Fluency Tests

Table 4: The most common testing methods in practice (based on Mohai, et al., 2016 pp. 44)

Since the dissertation focuses on the detectable changes in the EFs at the behavioral level, methods that are normally applied in neuropsychological research are not listed in the table. However, Mohai and colleagues (2016) point it out that the substitution of paper-and-pen tests with digitalized ones has long been debated due to the contradictiuous results in terms of the reliability of the test outcomes, they emphasize their applicability with prudent preliminary testing.

1.3.4.1 Measuring executive functions with fluency tests

Fluency tests are well-known, widely-accepted and applied measurement devices in clinical practice and research. In these tests the generation of random, unique and fast responses are evaluated with the requirement of minimal repetition in a restricted time frame. These procedural restrictions guarantee the application of a self-developed strategy in meeting the requirements. Despite the widespread use of these tests it is still unclear what EF components impact performance the most. Working memory is heavily involved in keeping the instructions and the previously generated words in mind. Inhibition is essential in suppressing irrelevant responses and repetition, while switching ability is challenged in modifying the search criteria (Shao et al., 2014).

Fluency is generally measured by two types of tests: verbal and nonverbal (graphics) tests (Diamond, 2013). However, sustained activation of attention is highly challenged during the generation of novel ideas in any fluency tasks, there is only partial overlap with regard to the underlying mechanisms in verbal fluency tests (when the emphasis is on selection) and in graphic tests (when the emphasis is on creation). The moderate correlation between the two test types also supports this assumption (Carone et al., 2007).

1.3.4.2 Verbal fluency test types

Verbal fluency is essential for proper social functioning and communication. Due to the involvement of the executive functions verbal fluency tests are often used in clinical research to support results of individuals diagnosed with attention-deficit/hyperactivity disorder (ADHD), Alzheimer's or Parkinson's disease or Autism Spectrum Disorder. These results might also give insight into individuals' lexical knowledge and lexical retrieval ability. Even though verbal fluency tests are quite similar regarding their construction and they are applied to measure both verbal abilities (semantic retrieval or the size of the lexicon) and executive functions (processing speed and working memory), test results cannot be interpreted unequivocally. Consequently, the underlying cognitive mechanisms and related neural circuits also remain unclear (Shao et al., 2014).

The classic verbal generation task has two variants: semantic and phonemic fluency tests. In semantic (category) fluency tests participants are asked to produce as many words as they are able to that fit into a certain semantic category within 60 seconds. Apart from the standard test procedures many other test recording techniques are known, mainly differing in the given time frame or the mode (oral or written) of the tests. In the semantic fluency condition the retrieval of lexical items is more automatic and lifelike since the items are semantically related. Semantic fluency depends on the knowledge stored in the semantic memory and on the individual's retrieval strategies; hence associative thinking is more typical in this type of fluency task (Luo et al., 2010). This task type relies on response inhibition (inhibition of automatic responses) since lexical items that do not belong to the same category specified by the task should be inhibited.

In the standard phonemic fluency test, COWAT, (Controlled Oral Word Association Test) participants are asked to produce as many words beginning with F, A and S as they are able to within 60 seconds (Troyer et al., 1998). After testing the basic indicators (number of generated words and perseverations), clusters (with at least two words), shifts between clusters, long words and unique words can be assessed to see the underlying cognitive mechanisms. Clustering and shifting have the same relevance in the analysis however, the number of words and their length might reveal the level of language proficiency (Kavé et al., 2008). Letter fluency tests draw on a mental process rarely done in everyday communication and for this reason the operation of some underlying cognitive processes, like self-monitoring, attention, processing speed, language processing and working memory is shared in both test types, that of others, like search strategies and memory is not. In the phonemic fluency condition results depend on the search and retrieval processes within orthographic or phonological networks, hence a less automatic and more abstract way of thinking is needed. Higher order cognitive functions are more involved in this test type since the participant has to suppress semantically or associatively similar distractors which results in the application of novel word retrieval strategies. Therefore, phonemic fluency relies on the level interference suppression (Robinson et al., 2012; Biesbroek et al., 2015, Leite et al., 2016).

In sum, semantic fluency test scores are generally accepted to draw more strongly on verbal ability and letter fluency test scores on executive ability, so individuals with deficit in lexical access or executive abilities will perform poorly in these tests (Shao et al., 2014).

1.4 The mental lexicon

Research on the bilingual mental lexicon focuses on the organization and the links of conceptual and form representations in the semantic memory. Navracsics (2007a) points it out that no model on language acquisition can be conceptualized without accepting the existence of the mental lexicon.

The long term memory is divided into two different subsystems: the declarative and the procedural memory. Unlike procedural memory, the declarative memory contains explicit knowledge that is consciously available, made up of words, while the procedural memory system is responsible for proper use of grammar, skills and habits. While natural language learning basically relies on the procedural memory, the declarative memory has a crucial role in instructed language learning. The mental lexicon is not only responsible for the storage of lexical items, but for those features as well, that make their use possible like pronunciation, written fom, word type, referential and extentional meaning (Gósy, 2005; Navracsics, 2007a; de Groot & Dukes, 2011; de Groot, 2013; Lukács et al, 2014). The structure of the mental lexicon is a constantly changing dynamic system of lexical items since it is influenced by different factors like age and context. For this reason the number of lexical items stored in it, can only be estimated. It comprises activated, passive and recently activated words. Items in the active vocabulary might become passive or those in the passive vocabulary might become active. The durability of this shift is up to the context (Gósy, 2005; Navracsics, 2007a). The most recent theoretical model on the organization of the mental lexion holds that lexical items are stored in a network-like system in which semantic units are interconnected. The two ends of a connection are called nodes to which many other connections might lead to. Language structure (whether it is an agglutinative or non-agglutinative language) has influence on that of the mental lexicon. The analytic approach holds that only base forms of words are stored in the mental lexicon and then their morphologically more complex forms are put together. Conversely, the holistic approach posits that these forms are stored as separate units. A third view holds that frequency of occurrence in case of suffixes might have influence on the way a lexical unit is stored (Gósy, 2005).

Models describing the bilingual lexicon take stand for distributed or unitary representations of the bilingual's two languages, whether they are completely or partially shared. While distribution models posit that certain conceptual representations are language-specific, unitary models assume the opposite. Another difference between the models is the assumed strengths and directions of the links. Finally, some of them also hold that change in L2 proficiency has impact on the links between forms and conceptual representations (de Groot, 2013). Moreover, de Groot (2013) also claims that the organization of the lexico-semantic memory is influenced by many different factors such as context, level of L2 proficiency, learning strategy and word type.

1.4.1 Bilingual lexical access

Bilingual models on lexical access, regardless of focusing on production or recognition, describe the selection process as language-specific or non-specific. Models considering bilingual lexical selection as language-specific, posit that only target language lexical items are activated, therefore the selection process is similar to that of monolinguals'. However, non-specific selection models hold that during processing the bilingual's all languages are involved (de Groot, 2011). In general it can be stated that more models have been formulated on bilingual language perception (comprehension) than on production (since this latter one has only received limited attention). Basically, speech production covers a different (top-down) process in comparison to language perception, since it refers to the speaker's deliberate decision on the entire speech act which involves the content, the proper register and the language. Most production-related studies regardless whether dealing with monolingual or bilingual speech production have focused on word production.

1.4.2 Factors influencing bilingual lexical retrieval

Since the mental lexicon is described as a network-like system in which semantic units are interconnected, tests that require the retrieval of semantically related words are more viable. In order to reveal the structure of the mental lexicon, semantic (category) fluency tests are frequently applied. In contrast, in phonemic fluency tests, that are applied in this study, word-retrieval is hardly based on everyday practice. However, it has to be considered that, in the written test format, the participants might look back at the previous word, therefore the retrieval of semantically-related upcoming words is more likely. For this reason the factors that might influence the retrieval processes based on semantics are detailed in the following sections.

Lexical retrieval follows a hectic pattern: the process is automatic until the retrieval of the first few words, then a longer pause occurs. As language use is operationalized in alignment with the law of minimum, the most easily accessible words are produced foremost. Word frequency is an influential factor that has impact on lexical processing in both perception and production. Frequent words are short and can be retrieved more quickly and smoothly in comparison to rare and longer ones, therefore the number of retrieval errors are also limited. Moreover, recently correlation has been found between the effort neccessary to retrieve a word and the exact time of its acquisition: the more frequent the word, the earlier it was mastered. It has recently been evidenced that word frequency influences not only the time of retrieval but that of production as well (Gósy, 2005).

Word type is another factor that has influence on retrival. With regard to word types, 50-60 percent of the generated words are nouns, 15-20 percent of them are verbs, in contrast, numbers, modals and adversb are quite rare. (Gósy, 2005; Navracsics, 2007b; Lukács et al., 2014). It is widely accepted that process and acquisition of abstract concepts are more difficult, hence are accomplished at a later age in comparison to those of concrete nouns. Imagery is an another indicator to define concreteness or abstractness of lexical items. The more abstract meaning a word has, the more complex verbal paraphrase its definition requires. Conversely, concrete words have unequivocal images, therefore they can be defined with ease. Absract words are also assumed to have a motority effect, however it might be triggered rather on the basis of their meaning, than on their abstractness (Lukács et al., 2014).

Regarding the L2, paired-associate studies reveal that learning and recalling foreign language equivalents of abstract words in the L1 is more difficult than that of concrete words. Besides correctness, context availability refers to the ease of availability of certain contexts in which the word normally occurs. Moreover, prior semantic or phonological knowledge stored in the long-term memory might also ease the intake of new words regardless of their concreteness (Groot, 2011).

If the participant is able to apply a retrieval strategy, reaction time declines and the sequence is repeated. Semantic and phonetic-phonological retrieval strategies are the most frequently applied ones. The number of syllables or the beginning or ending syllable in a stimulus word might also influence access. Apart from strategy use individual speed of lexical access and the size of the mental lexicon might also impact retrieval (Gósy, 2005).

1.5 Research on bilingualism

If experience can cause changes in the brain at structural and functional levels, life experience, such as intensive language use, might be one such experience. However, findings related to bilingualism are getting more and more controversial over the years. Most of the studies conducted on bilingualism investigate differences between bilingual and monolingual groups of participants with regard to their verbal or non-verbal abilities. Bilinguals are often reported to perform poorly in tasks measuring verbal abilities like lexical retrieval, although the explanation given are also various. Conversely, they are often reported to outperform their monolingual counterparts in tests measuring executive functions, specifically inhibitory control abilities. In the following sections these findings and explanations are detailed.

Language perception and production are natural ongoing processes that are never initiated, cut off or switched off, hence escorting each stage of human life. If two languages are acquired at a time, the task load is multiplied, because while one of them is being used, the other language is also active to a certain extent. This task load might be responsible for slowing down lexical retrieval and resulting in increased cognitive control (specifically inhibitory control abilities) that is assumed to be transferrable to other non-verbal domains (Bialystok, 2017; Paap et al., 2018). The assumption of having better inhibitory control abilities relies on Green’s (1998) inhibitory control model (IC). It highlights that the joint activation of the bilinguals’ languages contributes to the suppression of the non-targeted language equivalents while challenging their attentional system. At the same time, bilinguals seem to underperform monolinguals in tests of verbal abilities. Research on verbal abilities usually find smaller vocabularies for bilinguals in each age range in the picture naming task and the verbal fluency task

(Bialystok et al, 2010). Although, their weaker results on lexical retrieval (related mainly to the semantic memory) are explained by various reasons like the number of competing languages and words, the joint activation of the languages, the less time recruited to either language, context or the many other individual factors that might have influence on test outcomes. In contrast, bilinguals are often found to perform better in the phonemic fluency tests, than monolinguals since they rely more on executive functions (Bialystok et al., 2010; Abutalebi & Green, 2007; Bialystok & Poarch, 2014; Sullivan et al., 2017; Marsh et al., 2019).

Not surprisingly in bilingualism research, inhibitory control has been investigated the most extensively among all executive functions, although the results are quite mixed. Studies investigating inhibitory control of individuals apply tasks in which resistance to interference from distraction (a conflict) is needed. These tasks are mainly visual tasks, like ANT, Flanker, Simon-effect, Embedded figures task or semi-visual, like the Stroop task. The general finding is that when a conflict comes, reaction times increase which is the indicator of lower processing speed. Bilingual children and adults were often found to outperform their monolingual counterparts in tasks when interference suppression was needed (Kroll-Bialystok, 2013). In the previous years, bilingualism research focused on the existence or non-existence of the so-called bilingual advantage in inhibitory control.

In 2010, Bialystok applied an integrative test battery to investigate 6-year-old monolingual (N=25) and bilingual children's (N=26) executive control differences. The children took part in the global-local and the TMT test (both are visual tests), a receptive vocabulary test and a category fluency task. While the global-local test is used for measuring inhibition of attention in case of salient cues, TMT is applied for diagnosing motor speed, attentional control deficits and working memory functioning. While the monolingual and bilingual participants performed similarly on tests of receptive vocabulary and category fluency, bilingual children completed the executive function tasks in a shorter period even in the tasks lacking explicit conflict (in which no inhibitory control involvement is required). Bialystok and colleagues conclude that inhibitory control does not explain bilingual children's better performance. In contrast, Paap and Greenberg (2013) find no evidence for bilinguals' better executive processing. They conducted three studies in which they compared monolingual and bilingual adults carefully matched on parents' educational level and participants' L2 proficiency in a series of non-linguistic interference tasks (Simon, Flanker, color-shape, antisaccade). They did not find any evidence confirming the existence of any cognitive enhancement in bilinguals.

However, the often contradictory results encouraged researchers to conduct further studies. Although, they often apply different test batteries with the involvement of populations that are not controlled for certain factors. Researchers started to control participants for specific factors to explain the confounding results. In their study, Pelham and Abrams (2013) controlled the participants for their age of language acquisition. They measured participants' inhibitory control via the ANT task and their lexical retrieval via a picture naming task. They compared a group of monolinguals and a group of early and late bilinguals' performance to one another. Their results reveal that both bilingual groups perform equally (well) in the executive function task compared to monolinguals. Monolinguals' performance in the lexical retrieval task was significantly better than that of bilinguals, they also had the same deficit in the lexical retrieval task in their most dominant language.Surprisingly, the two bilingual groups did not differ between the number of frequent and rare words. Since participants were controlled for educational level and age, these factors were not defined as the cause of the results. The authors claimed that habitual use might have caused these results, rather than age of acquisition. However, results of the (semantic) lexical retrieval test are in line with Grosjean's (2016) Complementary Principle. Since bilinguals use their languages in different domains there can be no complete overlap between the lexical items of their languages.Kalia and colleagues (2014) do not confirm early bilinguals' superiority in EF tests. In their study, they compare monolinguals and two groups of self-reported bilinguals (ages 18 to 22) of different nationalities. Bilinguals were classified as early or late bilinguals regarding the age of acquisition of the second language (before or after the sixth year of life). Their executive functions were assessed by the ACNNT test (requiring the working memory in holding the rules of the task and inhibitory control). The participants' performance was assessed in terms of accuracy, inhibition, monitoring and switching. Their English receptive and expressive vocabulary and phonological awareness were also tested. The authors found no significant differences between the early bilinguals' and the monolinguals' performances in terms of inhibition (the number of perseverations) and switching abilities (mean RT). However, they reported a deficit in late bilinguals' accuracy in the tests in comparison to the other groups. Shishkin & Ecke (2018) investigated Russian and English (N=43) bilinguals (Russian immigrants), who spent similar amounts of time (14 years) in L1 environment (the USA), in the verbal fluency (semantic and phonemic) and Stroop task. At the time of the tests, younger participants were English dominant using English in twice as many contexts (11) as the other group and having been enrolled in English formal education for nearly 11 years. In contrast, the older participants were Russian dominant bilinguals, with less time spent in formal education (2.18 years) and fewer possible contexts to use English (5.59). They found that older immigrants' lexical retrieval was less fluent in their L2 and they remained dominant in their L1. Conversely, younger immigrants became balanced in their L2 proficiency. However, no significant difference was found between the groups in terms of controlling interference. The authors' conclusion was that younger participants might have had qualitatively different language experiences in comparison to their older counterparts. Moreover, they also assumed that it was not the balanced bilingual status that was the reason for their better control ability, but the stability of the language systems. Kousaie and colleagues (2014) came to similar conclusion in terms of the impact of language context on research outcomes. In their study, they investigated verbal fluency and executive functioning (working memory and cognitive flexibility) of highly proficient younger and older monolingual Anglophones, monolingual Francophones and French-English bilinguals, who were controlled for age, education, handedness, and general cognitive functions. They found contrasting results in the two tasks measuring interference suppression (Stroop and Simon), from which they concluded that the task type (verbal or non-verbal) might have influenced the test outcomes. Bilinguals' results on the verbal fluency tests were also inconsistent with the literature (monolingual Anglophones outperformed the other two groups). Overall, their results on executive functions did not confirm in bilinguals the enhanced cognitive functions since their findings were not consistent among the different tests.

Some researchers assumed that the size of bilinguals' vocabulary might have had an influence on the research outcomes. Luo and colleagues (2010) compared the performance of two groups of bilingual (low and high-vocabulary bilinguals) and a group of monolingual adults in verbal fluency test. Their receptive and expressive vocabulary differed, though their self-rated English language proficiency did not. Bilinguals and monolinguals performed equally in the category fluency test, but the high-vocabulary bilinguals outperformed both monolinguals and low-vocabulary bilinguals in the letter fluency criterion. The applied time-course analysis revealed better executive control abilities for both bilingual groups. The authors emphasize the influence of both language proficiency (vocabulary size) and executive functioning on the test outcomes of high-vocabulary bilinguals. They conclude that if the task is mainly based on lexical retrieval and the involvement of executive functions is less remarkable (like in a semantic fluency test) bilinguals' test outcomes are poorer than monolinguals. In relation to vocabulary size, Sullivan and colleagues (2017) assumed that frequency of words might have a role in the negative test outcomes on semantic fluency tests. They investigated monolinguals', bilinguals' and trilinguals' (N=200) performance in a picture naming task. The participants were all controlled for their receptive vocabulary size, ages and proficiency levels of additional languages. They all reported high proficiency level in each language. The authors found that low-frequency words were retrieved more slowly with a higher cost for the bilingual and trilingual group in terms of accuracy. However, results of each group demonstrated word frequency effects (longer reaction times for less frequent words). They concluded that it was the constant competition between the two languages for activation that was proved for the negative test outcomes, not the age of acquisition. Pino Escobar and colleagues (2018) compared monolingual and bilingual school-aged children (eight-year-olds) on tests measuring verbal fluency (phonemic and semantic) and inhibitory control abilities. The groups were matched in terms of vocabulary size and self-reported language use (provided by parents). Results reported on bilinguals' outperforming their monolingual counterparts in both verbal fluency tests. However, their performances in the inhibitory control tests were similar. The authors conclude that the involvement of inhibitory control processes in lexical retrieval is crucial, and so is the size of receptive vocabulary. However, Friesen and colleagues (2021) have come to a different conclusion. They investigated the linguistic and cognitive underpinnings in monolingual and bilingual children's performance in tests of phonemic fluency, English receptive vocabulary, word reading, fluid intelligence and working memory. Results revealed monolinguals' better performance in the receptive vocabulary test, but no difference was found between the groups in terms of verbal fluency tests. The monolinguals' performance was impacted by the size of their receptive vocabulary. For the bilinguals, both the size of their receptive vocabulary and their fluid intelligence had an impact on their performance.

The authors conclude that bilingual children seem to recruit additional cognitive resources to meet task demands required by fluency tests.

As opposed to Luo and colleagues (2010), Shao and colleagues (2014) find different reasons behind better results related to inhibitory control. Their study is the first that aimed at tapping the possible relations among different executive functions and in which the results correlate. They measured Dutch elderly (aged 60-89 monolingual participants' verbal fluency in the same way as Luo and colleagues (2010), and they also tested inhibition and shifting ability, but separately. They claim that instead of inhibition, updating predicted the total number of words produced in both test types. In the same vein, Marsh and colleagues (2019) assume different strategies beyond bilinguals' higher phonemic fluency outcomes. They re-analyzed the phonemic fluency test results of a longitudinal study on 200 young and old bilinguals. The participants were originally controlled for all the variables that could have influenced research outcomes (self-reported language proficiency, socio-economic status, patterns and context of language use, cultural differences, and typological similarity between the languages). Their findings confirm bilinguals' higher results in the phonemic fluency test. In addition, they also found evidence for the effect of bilingualism on switching and clustering abilities. Since results on shifting and clustering ability did not alter during the 25-year-period, they could verify Troyer's claim (1997) that age did not have an impact on test outcomes. In their study, Patra and colleagues (2019) aimed at measuring phonemic and semantic fluency performance and executive control (inhibitory control, mental flexibility and working memory separately). They compared 25 Bengali-English bilingual and 25 English monolingual adults. They were controlled for receptive vocabulary, language combination, proficiency, age, education, non-verbal intelligence). The authors applied clustering and switching while evaluating the fluency performance. Since their results revealed bilinguals' better performance on the letter fluency task, the authors drew the conclusion about their superior executive functioning. They also found larger cluster sizes in the letter fluency condition for bilinguals. In contrast, they did not expect but found equal switching values for the experimental and the control group. They concluded that bilinguals did not use switching as a strategy. However, they found significant differences between the inhibitory control measures only in favour of bilinguals, but no difference between measures of working memory functioning.

However, more recently, research on bilingualism has taken a new direction. Instead of inhibitory control processes, domain-general (top-down) attention is assumed to be responsible for bilinguals' better performance in tests measuring cognitive abilities (Bialystok, 2017). This selection mechanism is assumed to be responsible for the goal-driven focus that makes bilinguals able to ignore distracting information in tests that involve some explicit conflict like in Stroop. In their study, Friesen and colleagues (2015) tested bilingual (N=56) and monolingual (N=53) young adults' performance in two different visual attention tasks (feature-search and conjunction-search). The difference between the two test types is the number of dimensions on which participants need to focus. In case of the feature-search task type, participants need to consider only one dimension (a different colour), while in case of the conjunction-search test they have to focus on more dimensions (colour and shape) at the same time. The task load is higher in case of the conjunction-search task type. In this particular study (Friesen et al., 2015), bilinguals outperformed monolinguals in the conjunction-search test condition, while no difference was detected between them in the feature-search test condition. Furthermore, no explicit conflict is actually involved in these test types. The authors' conclusion was that bilinguals are accustomed to stimuli-rich contexts and this fact might have contributed to the test outcomes. As opposed to these results, Ratiu and colleagues (2017) found no marginal differences between their two groups' performance in either test condition. Paap and colleagues (2018) confirmed these results. They aimed to closely replicate Friesen and colleagues' (2015) research on bilinguals. For this reason, they recruited college students (N=141) and tested them in a conjunctive visual search task and an ambiguous figures task. Instead of categorizing participants as bilinguals and monolinguals, they gathered data on their proficiency level of the L1 and L2 and the ratio of the used languages. The authors found no effects of bilingualism in any of the tests. They also proposed the necessity of rigorous specification of the revised hypothesis (Bialystok et al., 2009; Bialystok, 2017) referring to its various descriptions in different studies (executive attention, selective attention or disengagement of attention). For this reason, they introduce the term attentional control and define it as a construct being crucial for bilingual language control. However, Paap and colleagues' (2015) claim is confirmed by Gärtner and Strobel (2021), who also emphasize the careful reconsideration of the applied tasks as conflict resolution might rely on either selection or inhibition.

The possible impact of bilingualism on working memory functioning and cognitive flexibility is even less clear. Enhancement in working memory functioning as a result of an extensive use of two or more languages has recently been recognized, though the findings are mixed. First, Bialystok and colleagues rejected that working memory might have an impact on L2 learning (Bialystok et al, 2009). Years later, they claimed that working memory might be enhanced under specific circumstances (Bialystok et al, 2012). However, the tasks, in which better working memory functioning has been detected, are visual tasks. In a meta-analysis of 79 studies, Linck and colleagues (2014) find that L2 proficiency and processing are positively associated with working memory. They also claim that the relationship between L2 proficiency and working memory functioning is so robust, that when L2 is involved in a working memory task, results also involve both. For this reason, the involvement of the L1 in a working memory task seems to be grounded if clear estimation of working memory functioning is required. They also find that verbal working memory measures show higher correlation with L2 outcomes than non-verbal measures. They claim that these results are more in line with Baddeley and Hitch's (1994) model on the working memory system (phonological loop), than with current models, like Engle and colleagues' (1999). They also claim that education, L1 abilities, length of L2 exposure might have an impact on the data. Calvo and colleagues (2016) posit that only specific aspects of working memory functioning might be influenced by bilingualism. Antón and colleagues (2019) investigated 180 young Spanish adults of whom 90 were Basque-Spanish bilinguals, controlled for general proficiency, age of acquisition, age, socio-economic status and IQ. They were tested in Simon, Stroop, Flanker, numeral Stroop, Corsi, Corsi-inverse, digit span and inverse digit span tasks. They conclude that bilingualism does not have a positive influence on general executive functions when the participants are carefully matched. However, they also claimed that there is no effect of bilingualism in the easiest working memory tasks found, improvement is detected only when the tasks require active processing and retrieval. The authors call for the specification of the extent (i.e. subsystems) at which bilingualism might have impact on the executive functions. They also emphasize that other factors (socio-economic status and memory abilities) might also have an influential role in the research outcomes in studies when participants are unmatched.

The inconsistent research findings related to bilingualism has led to various meta-analyses on the factors that might have impacted the outcomes. In their study Paap and colleagues (2015) listed vast many reasons why the existence of the ‘bilingual advantage’ (BA) can be contested. They claim that studies supporting the existence of BA were conducted on small sample sizes or with participants of unmatched socio-economic backgrounds. The application of the often inadequate statistical analyses might also have contributed to low replicability of the same test outcomes. Another problem is that the tasks normally applied for tapping inhibitory control do not show convergent validity. Publication bias is another sensitive issue related to bilingual research. Since mainly studies with positive outcomes are published, those with negative, mixed or null results have not been welcome. Even neuroimaging studies seem to result in inconsistent findings. Although they often report on bilinguals’ different neural processing, it does not necessarily mean that the use of the two languages is accompanied by enhanced executive functions at the behavioural level. They conclude that if BA in EF exist, they might be restricted only to particular components (Paap et al., 2015). In their meta-analysis, Lehtonen and colleagues (2018) come to the same conclusions. They reviewed studies on bilingual adults’ EFs including inhibitory control, monitoring, shifting, working memory, attention and verbal fluency. They investigated 152 studies, but found no advantages related to bilingualism. They also considered the publication bias of the investigated studies in their analysis. The authors confirm Paap and colleagues’ results, that is, studies reporting on positive outcomes were published and positive research outcomes were conducted on small samples. However, they also find that the difference between the results of monolinguals and bilinguals is smaller in non-verbal tasks. As opposed to Troyer (2000), they found no supporting evidence for the superiority of older participants in cognitive test outcomes. The same was stated in case of language similarity: no evidence of its role in domains of EF has been revealed. Finally, Gunnerud and colleagues (2020) in their meta-analysis investigated the cognitive advantages of bilingualism. Although they carefully controlled for variables that are considered as influential in the scientific literature, they report on remarkable differences only in case of switching tasks performed by middle-class children. They conclude that managing attention between different languages does not result in better executive functioning. However, they claim that L1 and L2 language skills and exposure to L2 seem to have impact on research outcomes, hence they have to be considered for future research.

They also suggest the consideration of bilingualism as a continuous variable, because in that way individual developmental trajectories can also be taken into account.

1.6 Hypotheses

As it is shown by the leading literature reviewed in the introduction, there is still no consensus among researchers on the possible outcomes of language learning and this is partly due to its complex nature. The available definition of certain phenomena and the description of attributed cognitive correlates raise further scientific questions even from distinct fields of research. Most studies apply quantitative measures that are largely based on the results of tests revealing only limited amount of information on participants. For this reason a multidimensional approach is applied in this study to investigate a phenomenon from a different perspective with different methods. This mixed methods research aims at revealing whether systematic and regular exposure to a second language in instructed settings (CLIL) may result in qualitatively different mental operations (higher level of flexibilityof thoughts and attentional control) which can be detected via both quantitative and qualitative methods. In accordance with these research goals, the following hypotheses have been outlined:

H1: CLIL learners perform significantly better in the selective attention test compared to their peers instructed according to the general curriculum.

H2: There is a significant difference between the total number of words generated by CLIL learners performing English phonemic fluency tests as compared to the controls.

H3: There is a significant difference between the total number of words generated by CLIL learners performing Hungarian phonemic fluency tests as compared to the controls.

H4: There is a significant difference between the cluster sizes produced in the English phonemic fluency tests by the CLIL and control group participants.

H5: There is a significant difference between the cluster sizes created in the Hungarian phonemic fluency test by the CLIL group and the control group participants.

H6: CLIL learners create significantly more task-discrepant clusters than the control group participants in the English phonemic fluency test.

H7: There is a significant difference between the number of cluster switches created by the two groups in the English phonemic fluency test.

H8: There is a significant difference in terms of the distributions of the different word classes between the two groups.

CHAPTER 2. Research methods

In this chapter the research design, the participants, the context and the ethical parameters are described.

2.1. Methods

Sántha (2015) claims that human development as a dynamic and multifactorial system cannot be completely approached from direct mathematical perspectives in social sciences. Lowie and Verspoor (2019) hold the same concept in the field of applied lingusitics suggesting that both group studies and individual case studies seem necessary to reveal the individual (internal) differences and context-dependent (external) factors that contribute to the nonlinearity of human development. They highlight that the individual's development is hardly comparable to any other's, due to the many numerous unstable factors that might have potential effects on it. Furthermore, the constant interaction of these agents results in completely different learning trajectories in individuals. For this reason it is quite difficult to find participants for a research project who are exactly at the same level in all relevant respects (Lowie & Verspoor, 2019).

The combination of qualitative and quantitative methods lays foundation for this multidimensionality in research. The 'fine tuning' of quantitative and qualitative methods (Mixed Methods) is more than the simple use of two or more different (qualitative or quantitative) methods within the same study. It refers to the systematic mixing of qualitative and quantitative data and methods. Research based on Mixed Methods combines the variable orientation and longitudinal approach of quantitative studies and the case-orientedness of qualitative studies with the aim of in-depth investigation (Sántha, 2015).

Multidimensionality in Mixed Methods studies can be ensured by the application of one type of methodological triangulation (between methods) that is the combination of a quantitative and a qualitative method. Interpeting triangulation as a multidimensional approach to studies goes beyond the classical notion that triangulation is merely a criterion of validity, since in addition to methodological triangulation, that of theoretical, personal and data can also be effectuated (Sántha, 2015). In this particular research the ’between methods’ type of methodological triangulation was implemented.

Many methodological models are differentiated in terms of the construction of studies based on Mixed Methods. These models overlap to a large extent. The study design used in data collection is Creswell’s Sequential Explanatory Design (CSED).

- In the CSED, sequenceality is manifested by a quantitative large-sample study followed by a qualitative small-sample study (Creswell, 2012).
- In the research performed the qualitative method serves as the main method to investigate a subsample that is obtained from the quantitative test sample with the consideration of certain criteria (Quant →QUAL).
- The data received from structured interviews of selected participants further refine the results to serve deeper understanding.

2.2. Research design

<table>
<tr><th colspan="5">1ST PHASE:
quantitative data collection and analysis</th></tr>
<tr><th colspan="2">Participants</th><th>Aim</th><th>Instruments and statistical analysis</th><th>Variables</th></tr>
<tr><td colspan="2" rowspan="4">N=142
CLIL group (N=69)
Control group (N=73)</td><td>to collect data on SES and language use and exposure</td><td>SES questionnaire
LEAP-Q questionnaire</td><td>background data,
not involved in the statistical analysis</td></tr>
<tr><td rowspan="3">to select participants in each group with the highest and lowest test results</td><td>TESTS:
d2-R: test of selective attention;
(T-test, Mann-Whitney Test)</td><td>mean performance, number of processed target stimuli, omission errors, incorrect markings; concentration, accuracy</td></tr>
<tr><td rowspan="2">phonemic fluency tests:
English (FAS)
Hungarian (KTA)
(T-test; Mann-Whitney Test)</td><td>number of generated words
switching: cluster switches, hard switches
clustering: cluster size, phonological, semantic clusters
slight, strict clusters
perseverations, repetitions, errors</td></tr>
<tr><td>Additional variables:
number of abstract words, concrete words, nouns, verbs, adjectives, other word types</td></tr>
<tr><th colspan="5">2ND PHASE:
QUALITATIVE data collection and analysis</th></tr>
<tr><td>4 groups:
CLIL ’high’
Control ’high’
CLIL ’low’
Control ’low’</td><td>N=3 per each group</td><td>to collect complementary data in order to reveal complexity of thoughts</td><td>structured interview on L2 learning</td><td>Main codes:
aim of learning the L2, language teacher, lessons, activities, competences, development, do differently</td></tr>
</table>

Table 5: Research design

2.3. Participants and sampling criteria

Data were provided by schoolchildren of five primary schools, two of which offer a CLIL programme, in County Fejér, Hungary. Stratified (convenient) sampling technique was applied in which the base for stratification was the number of English lessons in the curriculum. Stratified sampling technique belongs to random sampling techniques in which each element of the population has the same chance of being included in the sample. I defined the layers based on the number of English lessons (Csíkos, 2009). This way two groups were defined: the experimental group (N=69, hereinafter referred to as CLIL group) and a control group (N=73, learning according to the general curriculum, with 3 or 5 English lessons per week). Participants were all eighth graders.

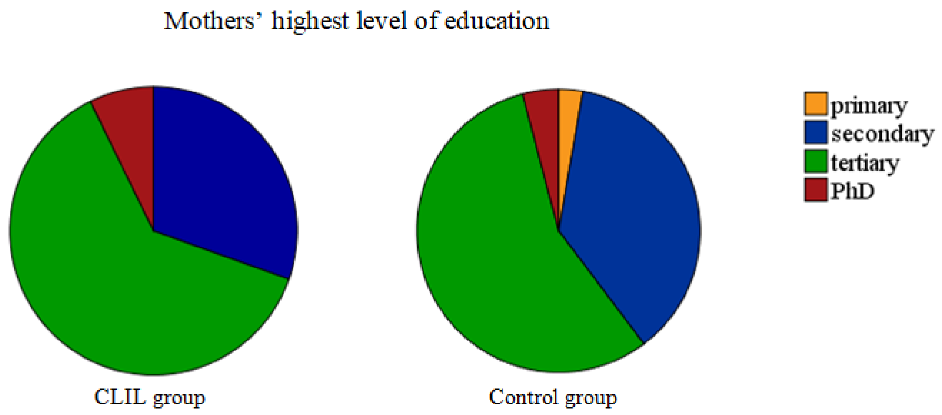
The aim of the application of the socio-economic questionnaire (SES) was to reveal basic background information about the participants such as age, gender, handedness, vision, parents' marital status and highest level of education 142 participants produced valid answers for further analysis of whom 69 were CLIL group and 73 control group participants. Considering the gender ratio, 44.9% of the participants identified themselves as males and 55.1% as females in the CLIL group. The mean age for them was 13.58 (*SD*=.49722) of whom 42% were 13 and 58% were 14 years old. 26.1% of the CLIL participants reported about corrected-to-normal vision, hence before the testing sessions all participants were asked to wear their glasses and at the time of test taking all participants acted as it was requested. Regarding handedness it can be concluded that 10.1% of the participants were left-handed. At the time of test taking 37.7% or the parents lived separately from the family. Mothers' highest level of education was reported as follows: 30.4% of the mothers finished secondary and 62.3% tertiary education and 7.2% of them had a doctoral degree. Fathers' highest educational level was reported as follows: 2.9% of them finished primary, 42% secondary and 52.2% tertiary education. Only 2.9% of them had a doctoral degree. According to the tests of normality (Kolmogorov-Smirnov and Shapiro-Wilk tests) there were no normal distribution in any of these parameters (p=.000, p=.000).

In terms of gender ratio, 50.7% of the participants identified themselves as males and 49% as females in the control group. The mean age for the control participants was 13.71 (*SD*=.51315) of whom 31.5% were 13, 65.8% were 14 and 2.7% were 15 years old). 37% of the control participants reported about corrected-to-normal vision, hence before the testing sessions all of them were asked to wear their glasses or contact lenses. At the time of test taking all participants involved acted as it was requested. Regarding handedness it can be concluded that 12% of the participants were left-handed. At the time of test taking 37% of the parents lived separately from their family. Mothers' highest level of education was reported as follows: 2.7% finished primary, 37% secondary and 56% tertiary education. 4.1% of the mothers had a doctoral degree. Fathers' highest level of education was reported as follows: 2.7% finished primary, 68.5% secondary and 26% tertiary education. Only 2.7% of them had a doctoral degree. According to the tests of normality (Kolmogorov-Smirnov and Shapiro-Wilk tests) there were no normal distribution in any of these parameters (p=.000, p=.000).

The Mann-Whitney test indicated that there was no statistically significant difference between mothers' highest levels of education in terms of the two groups (p=.168). In contrast, that of fathers' differed significantly (p=.004).

Based on the results of the Fisher-test it can be concluded that gender ratio (p=.507), state of vision (p=.207) and parents' marital status (p=1.000) are independent of the group (CLIL or control), that is the indicated difference is by chance.

The following pie charts report about the rates of the parents' highest level of education:



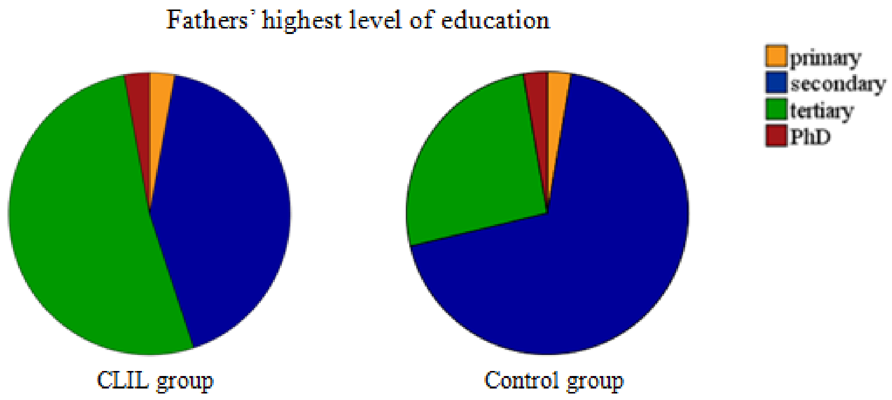


Figure 10: Parents' highest educational level

The data collection took place from September to December 2019, in five different schools in County Fejér with the written consent of the educational district director. Before the data collection, the participants' parents were informed about the aim of the research and asked for their written consent (see in Supplementary Material No. 1.)

The students took the tests and filled in the questionnaires in groups during a two-hour session (2x45 minutes) in their schools. Prior to the start of the session, they were informed about the nature of the investigation they could leave at any time without consequences. All testing sessions started at 9 o'clock in the morning in regular school time with a short explanation of the tasks. There was a 5-minute break after each task and a 15-minute break after the first session. The order of the tasks was fixed for two reasons: first, to adjust to the characteristics of the age group in terms of their need for diversity, and secondly, to reduce interference among the tasks. Only one test or questionnaire was placed on the participants' desks at a time. Participants were highly co-operative and there were no disturbing factors during the testing sessions.

After the evaluation, learners in either group (CLIL or control) achieving outstanding results in each test have been selected for interview questions. None of the respondents refused participation. The structured interviews were taken in the participants' schools and lasted for about 30 minutes.

2.4. Instruments, procedures, and data analysis

2.4.1. LEAP-Q

The Language Experience and Proficiency Questionnaire (Marian et al., 2007) is a standardized questionnaire for collecting self-reported data on any number of languages used by multilingual individuals to capture their language profiles. Although LEAP-Q is a self-report questionnaire, during the validation process, its developers found from moderate to high correlations between the proficiency levels and language measures for L2. LEAP-Q was originally constructed to be used for research settings with the involvement of mentally intact adults and adolescents. LEAP-Q was designed with the aim to cover as many factors that might have influence on bilingual experience as possible. For that reason, LEAP-Q contains questions related to language dominance (self-reported degree of foreign accent and the level of proficiency in the four language skills), language exposure (extent of language immersion and exposure), language preference and language background (milestones in language learning, age of onset). As Kaushanskaya and colleagues (2019, p. 1) point it out:

> ’At minimum, any work in bilingualism published today strives to include the following information: the ages at which the bilinguals’ two languages were acquired; the extent of exposure to the two languages currently…; and estimates of dominance and/or proficiency (subjective, objective, or both).’

In the study analysed in the dissertation, LEAP-Q was applied with the intention to gain additional information on participants’ language background that cannot be collected through direct testing. Moreover, the minimum output criteria declared in the corresponding decrees on language development specify and guarantee the language levels the learners of CLIL or general programmes need to achieve annually.

For the dissertation the paper-and-pen questionnaire had been translated and applied since the online revised (and translated) format was released years later. Since LEAP-Q is validated in its original format, the authors of the questionnaire do not encourage deletions, insertions, or other changes (order of questions or wording). Nevertheless, the authors’ recommendation is that additions need to be inserted at the end of the questionnaire (Marian et al., 2019). The only modification implemented in the questionnaire was a list of those activities that characterize the age group in the L1 and the L2 (reading, writing blogs, watching films etc.) and placed at the end of the questionnaire as recommended.

Participants completed the questionnaire in approximately 20 minutes. Extensive explanation was given to them before the completion of all questions. Most of them were easily understandable even for 13-15-year-olds, but if participants needed help, it was given individually and instantly.

2.4.2. Test d2-R

Test d2-R is a widely applied paper and pen test in clinical practice and research. It is used in the fields of neuropsychology, psychiatry, educational psychology, career counselling, work psychology, sports psychology as well as in the selection and maintenance of personnel for jobs and activities that require great responsibility and vigilance. Test d2-R is considered as a general performance test which measures concentration and the ability of sorting out irrelevant stimuli at the same time. For this reason, it cannot be regarded as a pure selective attention test, although it has the advantage of not confusing attentional performance with other abilities like counting. Due to its structure, the test can be used for all age groups between 9 and 60 years (Brickenkamp et al., 2010).

According to the test manual, the test battery meets the following criteria set by Westhoff and Hagemeister in 2005 (cited by Brickenkamp et al., 2010).

- oral and written instructions need to be short and to the point,
- target stimuli must be easily detectable,
- the opportunity for a short ’practice’ before real testing (to check understanding) is necessary to be given,
- elimination of ceiling effect should be ensured (the increased number of target stimuli in each row is the guarantee for this),
- the test paper must be suitable for individuals and groups either,
- objectivity, reliability, and validity must be guaranteed.

The test has 798 signs (’d’-s and ’p’) of which ’d’-s with two lines should be detected. Both lines can be seen at the top or at the bottom, or one line at the top and the bottom of the letter ’d’. Target stimuli are arranged in 14 rows with 57 signs in each. The rows are systematically repeated in the test (three rows constitute a block). A single block contains 171 signs from which 94 are irrelevant stimuli and 77 are target stimuli.

2.4.2.1. Testing session

Participants are first asked to complete the data necessary for identification and then get familiar with the test. If this preparation phase is over, participants start marking (crossing out) all signs (letter d-s with two lines) in the two practice tasks. This process can be done in the participants' own paces. In the introductory phase, the test administrator emphasizes that the direction of the 'crossing mark' is not of relevance, but the tracking move is necessary to be done from left to right in each row. The participants have 20 seconds for each row; therefore, the entire testing session lasts for 4 minutes and 40 seconds. The real testing phase is initiated with the announcement of 'Start'. (It is important to notice that understanding of the task can only be guaranteed if the style and wording of instructions are adjusted to the participants' age.) After the end of 20 seconds the test administrator says the following sentences: 'Stop. Next row'. The process continues till the end of the last 20 seconds when the test administrator finishes this way: 'Stop. Time is up. Put your pen down'. Although this seems a rather short period for testing attention, effective processing requires continuous and intensive concentration and immediate reaction while the frequency of stimuli is high (Brickenkamp et al., 2010).

2.4.2.2. Scoring criteria

The following scoring principles are applied: the first and the last rows are not considered and seen as trials, so actually 12 lines are assessed with 308 signs. Two error types can be detected: incorrect markings and omission errors. 'Incorrect marking' refers to a sign that has been marked erroneously. 'Omission error' is made if a sign should have been crossed out so it is left unmarked. If incorrect markings and omission errors are detected, they must be subtracted from the number of processed target stimuli to gain the performance value of concentration. Work pace (processed target stimuli) and accuracy (rate of errors) determine the effectiveness of task performance (the performance value). Finally, the raw scores are converted into standard scores with the help of normative data tables. This way results can be interpreted in comparisons to the performance of the members of the same age group. Due to the limited accuracy of the test, determining confidence intervals are necessary to reveal the extent to which the score obtained may deviate from the ’real’ score (Brickenkamp et al., 2010).

2.4.3. Phonemic fluency tests

Verbal fluency tests are applied in both clinical practice and research since they provide information about the word retrieval from the mental lexicon and those executive processes that are responsible for controlling it. In the standard phonemic (letter) fluency test participants are asked to produce as many words beginning with F, A and S for the English language test and K, T, A for the Hungarian language test as they can within 60 seconds for each letter. Participants are also asked to avoid mentioning geographical names or proper names (Troyer et al., 1998; Abwender et al, 2001; Tánczos et al, 2014).

Although, in clinical research mainly quantitative analysis is applied, Toyer and colleagues’ (1997) qualitative scoring system for verbal fluency emphasizes the importance of the underlying strategies in the retrieval process. For this reason, they initiated the application of a processed-based qualitative analytical approach that distinguishes clusters and switches as dissociable components of fluency performance and as signs of intentional strategy use. In accordance with these principles a similar scoring method is elaborated and applied in the investigation described in the dissertation to reveal different aspects of strategy use (Troyer et al., 1997; Mészáros, 2017).

2.4.3.1. Quantitative analysis – general rules

The following scoring principles were applied for both languages:

- 1 score was given for each correct word,
- Errors and repetitions were not included in the total. Errors include words that begin with a wrong letter, are proper nouns, geographical names or have suffixes. Two types of repetitions were differentiated. While perseverations are defined as the immediate appearance of the same word twice, repetitions are detected in case of later reappearance (Troyer, 2000).
- ’perseverations’ were detected if the same words appeared right after each other,
- words that were scattered again, were considered as ’repetitions’.

Basically, the number of generated words, the mean cluster size, and the number of switches are calculated (Troyer et al, 1997).

2.4.3.2. Hungarian language – scoring principles

As Hungarian is an agglutinative language, the application of a more elaborated analysis seemed necessary that was based on Mészáros and colleagues' (2011) work. Table 6 reveals the applied scoring principles based on their scoring system:

PHONEMIC FLUENCY TEST – general scoring rules			
No extra score (only 1) is given for their co-occurrence		**1 scores are given for each word with the following suffixes**	
conjugated or suffixed words in a row	(fa, fát)	-ság, - ség	(piros-pirosság)
preverb+verb	(kiáll, kinéz)	-itás	(naiv-naivitás)
noun with a diminutive suffix	(cica, cicácska)	-s (based on noun form)	(kuka-kukás)
suffix -ás, -és for noun formation	takarít-takarítás (except with a different meaning: áll-állás)	-atlan, -etlen, -talan, -telen	(hely-helytelen)
ó,ő ending	(fut-futó except it is differentiated as a headword in the dictionary: tanít-tanító)	-nyi	(pohár-pohárnyi)
i ending	(egyetem-egyetemi)	-l	(box-boxol)
-beli	(fajta-fajtabeli)	-ál	(analízis-analizál)
suffix s (based on adjective form)	(kék-kékes: except in case of a word with different meaning: dob-dobos)	-kodik	(okos-okoskodik)
permissive suffix: -hat, -het	(kap-kaphat)	-g	(csörög-csörren)
frequentative suffix	(lép-lépeget): except in case of a word with own meaning: tereget	-l	
causative structure	(olvas-olvastat; except with a different meaning: szoptat)		
-an, -en	(szép-szépen)		
-ul, -ül	(olasz-olaszul)		
-lag, -leg	(zene-zeneileg)		
-kor	(érkezés-érkezéskor)		
-ként, - képpen	(falu-falunként)		
-stul, - stül	(gyerek-gyerekestül)		
compound words	(kőház, kőút, kőkerítés): except they are headwords		

Table 6: Scoring principles (based on Mészáros, et al., 2011)

2.4.3.3. English language – scoring principles

As for the English language the same general principles and rules have been applied as in case of the Hungarian language.

Generally, a change in a word ending to produce a new word (with new meaning) that refers to a noun (e.g., “teach” and “teacher”) was considered acceptable and such instances are scored as two separate words. Homonyms of previous responses are accepted if the participant made the meanings clear. Slang, swear and commonly used foreign words are also scored as acceptable responses in accordance with Benton and colleagues’ work (1983).

2.4.3.4. Qualitative analysis

Clustering and switching are normally considered as signs of intentional strategy use in verbal fluency performance, although the degree of deliberateness behind these strategies is still not envinced (Abwender et al, 2001). However, it is generally accepted in the related literature that both strategies are necessary for the optimal performance in the fluency tests (Troyer, 2000).

2.4.3.4.1. Clustering

Clustering is defined as a highly automatic, strategic retrieval of words within phonemic or semantic subcategories (Abwender et al, 2001). The following scores have been calculated during the analysis. Firstly, cluster size, that is seen as a crucial indicator for the organization of the semantic memory and the effectiveness of word retrieval. In the dissertation the number of words actually produced is used to determine the size of the cluster, since the arbitrary application of n-1 in Troyer’s protocol (Troyer et al., 1997) does not seem to reflect the participants’ intention to produce a cluster. To compensate for this change and make the analysis even more detailed an alternative scoring system has been developed: the concept of distinguishing two types of clusters (slight and strict) that is based on the coding system applied by Mészáros (2017). Slight clusters consist of at least two words, while strict clusters contain three or more words. Troyer considered (1997) two or more neighbouring words with the same two initial letters (sell – self) or rhyming pairs (apple – ample, fight - flight) or homonyms (steal – steel) or words with only one differing middle vowel (fur – far; sit – seat) as a cluster. In case of overlapping clusters (in which one or some words are shared) the mutual words (*kukorica* in the example) are counted in both clusters (*kutya, kukorica, káposzta*) In accordance with Troyer's (1997) work repetitions, perseverations and errors are included in the clusters as signs of strategy use.

Although task-consistent clusters are more frequent and expected if the required criterium is met (phonemic clusters in the phonemic fluency test), in accordance with Abwender (2001) both task-consistent and task-discrepant clusters have been calculated in the dissertation in order to reveal signs of a more deliberate strategy use. Task-consistent clustering covers answers within the same phonemic or semantic criteria and task-discrepant clustering refers to inconsistent answers (e.g. a phonemic clustering in a semantic fluency task). Table 7 reports on the main clustering principles:

PHONEMIC FLUENCY TEST	
at least two words are generated in a cluster	
PHONOLOGICAL CLUSTER	**EXAMPLE**
same initial letters	expected (this is the task)
same 2 initial letters	spaletta-spagetti; sound-sour
rhyming pairs	kapál-kalapál; tree-three
same first and last letters	sír-sár-sör;it-fat
homonyms	só-show; foul-fowl
SEMANTIC CLUSTER	words that share the same meaning or are in the same semantic category

	Slight cluster	Strict cluster
Phonologial cluster (task-consistent)	foreign, formal túra, túró	fade, face, fake kenyér, kerámia, kereszt
Semantic cluster (task-discrepant)	sandwich, sausage kutya, kacsa	fancy, fast, funny, furious kifli, kenyér, kalács

Table 7: Clustering principles

2.4.3.4.2. Switching (shifting)

Troyer (1997) defines switching ability as the of frontal lobe functions that requires mental flexibility. In the dissertation I applied Abwender and colleagues' (2001) protocol on assessing clustering ability. They distinguish two types of switches ('cluster switching' and 'hard switching') to draw the attention to the probably different subserving mental mechanisms. They define cluster switches as the shifts between two adjacent or overlapping clusters and hard switches as the transition between a cluster and a single (non-clustered) word or between two single words. In contrast hard switches simply reflect the lack of clustering abiliy and represent the speed of information processing and mental productivity. In the dissertation this differentiation is adapted in order to gain a more subtle analysis. The two types of switches are coded separately and then the total number of raw switches is also calculated in the dissertation in accordance with the related literature (Troyer, 2000; Abwender, 2001). Table 8 reports on the general evaluation principles of fluency performance:

number of generated words	vocabulary size (Troyer, et al., 1997); working memory (updating) (Shao, et al., 2014)
irrelevant responses	working memory/inhibition (Shao et al, 2014);
perseveration	working memory/inhibition (Shao et al, 2014); sign of intentional strategy use (Troyer, et al., 1997)
repetitions	working memory/inhibition (Shao et al, 2014); sign of intentional strategy use (Troyer, et al., 1997); inhibition impairments (Delgado-Alvarez, et al., 2021)
less perseveration errors and more total words	better working memory performance (Amunts et al, 2020)
clustering	ability to access and use the word store (mental lexicon) (Troyer, et al., 1998); mental flexibility (Amunts, et al.,. 2020); episodic memory (Delgado-Alvarez, et al., 2021)
task-consistent clustering	clear sign of intentional strategy use (Troyer, et al., 1998)
task-discrepant clustering	real sign of intentional strategy use (as opposed to task-consistent clustering) (Abwender, 2001)
cluster size	organization of the semantic memory, effectiveness of word retrieval (Troyer et al., 1997); sign of mental flexibility, access words within a subcategory (Raskin, 1992)
switching (shifting)	lack of clustering ability (Abwender, et al., 2001); executive functioning (Delgado-Álvarez, et al., 2021); general executive functioning (Messer et al, 2018)
shifting (switching) between categories	mental flexibility (Amunts, et al, 2020; Troyer, et al., 1997); real sign of mental flexibility: greater degree of intentionality in strategy use (as opposed to other switching types (Abwender, et al., 2001)
general VF performance	role of attention in VF tasks (Amunts, et al, 2020); Delgado-Álvarez, et al., 2021); planning and cognitive flexibility (Leite, et al., 2016)
number of unique and academic words	lexical sophistication (Verspoor, et al., 2012)

Table 8: General evaluation principles of fluency performance

2.4.3.4.3. Additional aspects of fluency performance

To capture the differences in the lexicons of the two participating groups the number of unique and academic lexical items has been calculated. In accordance with Verspoor and colleagues' article (2012) lexical sophistication is measured by the number of unique words and academic vocabulary that were calculated at the group level.

In order to calculate the number of academic words and words that belong to different genres COCA (Corpus of Contemporary American English) and MNSZ2 (Magyar Nemzeti Szövegtár) databases were applied. The number of different word types (abstract and concrete nouns, verbs, adjectives and ’others’) was defined due to our assumption that their frequencies might reveal an imbalance between the different language levels of the two groups. Finally the topics of the generated words were also defined. In the dissertation unique words at the group level are defined as lexical items being mentioned only by one of the groups.

2.4.4. Structured interviews

According to Nádasi, (2000) the application of structured interviews is justified if the researcher aims to reveal participants’ views on a specific topic. Given that, learners at this age might not have explicit knowledge about those cognitive abilities that have been tested, interview questions must have been consciously created to meet our primary aim which was the detection of dissimilarities in CLIL and control group participants’ mindsets. Therefore, we sought to create questions that do not require any special expertise in any topics, but provide enough stimuli to make individuals’ specific approaches detectable. For this reason the topic of L2 learning have been chosen. This way information about the applied L2 methodology through which learners are really taught has been received. We have formulated some of the questions in such a way that they can be approached from different perspectives (e.g. What is the most important in learning English?) We assumed that participants with the highest tests results would provide more specific and diverse answers as the signs of mental flexibility.

Before the compilation of questions, a set of different criteria had been considered in line with Nádasi (2000). First, considering the age of participants (adolescents), questions had to reflect sincere interest while being motivating, straightforward and clearly worded. For this reason, the structured interview comprised open-ended questions in a special order among which core items were placed in the middle section. The order of questions was influenced by the following principles: sharp shifts between a) topics and b) viewpoints (internal-emotional or objective; imagination or reality) have been made. The application of qualitative interviews has some advantages: due to their exploratory nature, participants can express their opinion with high engagement and their thoughts can be analysed in the original context.

Regarding the number of questions (11) and the time constraints (approximately 30 minutes in a 45-minute lesson in regular school time), the written format was preferable, since it contributed to the indirect (self-paced) management of time without external intervention. The questions were written in the Hungarian language for two reasons: first, to eliminate anxiety and secondly, to avoid restricted language use that might have eventuated in less informative opinions. However, participants were given the opportunity to answer the interview questions in line with their language preference.

The text corpora consisted of two pages on average. The processing of the structured interviews was facilitated by the MAXQDA software.

For processing structured interview data, qualitative content analysis (QCA) with the combination of deductive (concept-driven) and inductive (data-driven) logic was applied (Kuckartz, 2019). During deductive content analysis the setting up of a code list (code frame) preceded the data analysis. This list of codes (categories) is required to be described precisely to ensure the reliability of the coding procedure which is only a part of the preparational phase (Kuckartz, 2019). In the study the reliability of coding was ensured by intracoding during which the author coded the interviews twice following the same coding system. The elapsed time between coding and re-coding was two weeks to eliminate remembrance. During re-coding there was no conceptual change, hence the reliability indicator of coding was 1.

The practice of 'a priori' coding is justified if the compilation of main codes is based solely on theory and/or the researcher's own experience (Sántha, 2015). After setting up the code frame, the next step was the inductive discovery of the deep layers of text segments. This time subcodes were searched first, and then a match between the meaning of texts and the subcodes. All text segments that are relevant to the research must be and were coded completely. The analytical process comprised the phases of summarizing, comparing, and contrasting data (Kuckartz, 2019). Both qualitative and quantitative techniques were applied in the data analysis. The qualitative analytical techniques contributed to the exploration of the different opinions of the groups on specific topics. In contrast, the application of quantitative techniques revealed the frequency of certain categories and subcategories.

2.4.4.1. Selection criteria for the structured interviews

For the qualitative content analysis 4 groups have been created: a high-performance CLIL group (referred to as CLIL ‘high’), a low-performance CLIL group (CLIL ‘low’), a high-performance control group (control ‘high’) and a low-performance control group (control ‘low’). In each group there are three participants. Participants who met the following criteria have been selected for the ‘high’ groups:

- regarding the D2-R test, participants had to achieve high performance level for attention (at least 110 standard points) coupled with high accuracy (at least 110 standard points). Any errors that may result from the measurement accuracy of the test have also been considered while constructing these scores. Considering the measurement error of the test, these participants achieved at least high or average performance in both parameters;
- regarding fluency tests, participants had to be among the five (positive) outliers listed by SPSS software in as many investigated parameters as possible.

Participants in the ‘low’ groups had to meet the following criteria:

- regarding the D2-R test, participants had to achieve low performance level for attention (maximum 94 standard points) coupled with low accuracy (at least 94 standard points). Any errors that may result from the measurement accuracy of the test have also been taken into account while constructing these scores. Considering the measurement error of the test, these participants achieved at least low performance in both parameters;
- regarding fluency tests, participants had to be among the five (negative) outliers listed by SPSS software in as many investigated parameters as possible.

CHAPTER 3. Data analysis and results

In this chapter, first quantitative data gained by the participants of the groups in the questionnaire and the different test types are presented and analysed.

3.1. LEAP-Q

3.1.1. CLIL group – language presence, language preference

Data was processed by SPSS software. **69 participants** reported Hungarian as their first language (language learnt from birth). With the exception of 3 participants everybody else reported English as their second language. The difference between the numbers of (mean) days spent in the first and second language environment is remarkable and confirms the dominance of the Hungarian language. While it is about 4551 in case of the first language, the same indicator has a different value (224 calendar days) in case of the second language. The same difference is observed when the numbers of days spent in the first and second language context are summed: the number of total calendar days in case of the first language is more than twenty times (277610) higher than that of the second language (13683). The shortest time period spent in first language environment is 1460 days, while the longest is 5110. Most CLIL learners (with the exception of 4 participants) spent 4551 calendar days in the first language environment until the day of data collection, however, this number is considered to be balanced (*SD*=697,24387) with a range of 3650 days (minimum: 1460, maximum: 5110). In this regard CLIL learners' second language features fundamental differences, with extreme values: a larger standard deviation (*SD*=802,07887) in comparison with the first language and an even larger range value of 4745 (minimum: 0 and maximum 4745).

In the questionnaire participants were asked to estimate the presence of the listed languages in their everyday lives. Based on the responses of the CLIL participants it can be stated that, the difference (nearly 17 percent) between the presence of their first ($\bar{x}$ =57,86 percent) and second ($\bar{x}$ =39,66 percent) languages in their lives is not noteworthy. In this case there were no extreme values, 0 percent minimum or 100 percent maximum, so the data set were more compact in each case (SD_{Hun} =13,811, SD_{Eng} =13,675).

The following section deals with CLIL participants language preferences in different contexts. In case of reading or listening to something of their interest, either Hungarian ($\bar{x}$=51,8382 percent) or English ($\bar{x}$=45,5000 percent) is chosen nearly at the same rate with balanced and close values of standard deviation (SD_{Hun}=24,85637, SD_{Eng}= 24,34485) and variance (σ^2_{Hun}= 617,839, σ^2_{Eng}=592,672) in both cases. The high values of variance reveal that there were participants who reported no presence of either the first or second language in terms of these activities. Based on these data, it can be stated that both languages are used actively in CLIL learners' everyday lives for activities like reading or listening.

The next question is aimed at revealing to what extent participants would choose either language for chatting with someone. Hungarian language would be chosen at a higher rate ($\bar{x}$=57,7647 percent) than the English language ($\bar{x}$=40,0000 percent). The high values of variance (σ^2_{Hun}= 602,451, σ^2_{Eng}=512,328) reveal that not only a few respondents produced extreme values. It can be concluded that CLIL learners seem to consciously choose the first language as opposed to the target language for activities like chatting.

3.1.2. CLIL group – L1 and L2 use

In the following sections the results of the applied methods (the Friedman-test and the Principal Components Analysis) are analyzed and reported in case of the estimated level of language skills and activities done in the first language. The Friedman test is applied when we aim to reveal the distribution of at least three ordinal variables within a group. Regarding the self-estimated level of first language skills, the result of the Friedman-test ($X^2(4)$=21,293, p=,000) reveals that there is no overall statistically significant difference between the mean ranks (listed below), that is, participants are on the same opinion. The lower the mean rank is the more important problem area it reflects, that is participants considered writing (2,41) as the most, and listening comprehension (3,39) as the least problematic area in their first language. In terms of the other skills, the following data have been revealed: speaking: 2.80, reading: 3.11, accent: 3.29.

The Principal Components Analysis is used when the number of variables from a single population is necessary to be reduced with the least data loss. PCA is used for revealing the system of relationships among different variables, hence the originally related variables. Twelve variables were included in the PCA (reading, listening to music, wqatching films, using the intrnet, texting, learning, online gaming, writing blogs, skype use, family chat, info search, watching TV). First, the KMO (Kaiser-Meyer-Olkin criterion is necessary to be met. This value guarantees that all the variables are adequate for the PCA. In this case KMO is 0,780 meaning that the model is adequate for the analysis, and it should yield reliable factors. Bartlett's Test of Sphericity reveals that the null hypothesis (there is no correlation between the variables) can be discarded since the level of significance is lower than 0,05 (p=,000) referring to the existing correlation between the variables, so they are adequate for the analysis. Overall, the adequate KMO values and the result of the Bartlett's Test justify the adequacy of variables for the factor analysis. Considering that, the level of self-estimated accent had the lowest communality value (h^2=0,003, revealing the weakest correlation with all other items) it was extracted from the list of variables, then the two tests were run again. This time the KMO value and the result of the Bartlett's Test, again, justified the adequacy of the variables for the analysis (KMO: 0.793, p=.000). Results show that while using their L1, CLIL learners regarded speaking, listening comprehension, writing and reading as equally important skills, among which listening comprehension was considered as the least problematic in their L1. Accent, the least remarkable factor, could not be included among the variables that make up the principal components of the model.

Regarding everyday activities in their first language, the result of the Friedman-test ($X^2(12)$=270,498, p=,000) reveals that there is no overall statistically significant difference between the mean ranks, that is, participants are on the same opinion. In this case, first language use is the least characterized using skype (3,34), writing blogs and texts (3,35) or playing online games with others (4,43). In contrast, the most frequent activities are family chat (10,06), learning (9,23) and chatting (9,13).

The condition for Principal Components Analysis, the adequate KMO (0,672) value and significance (p=0,000) level are both met. Overall, the adequate KMO values and the result of the Bartlett's Test justify the adequacy of variables for the factor analysis. Considering that, reading (h^2=0,064), listening to music (h^2=0,034), internet use (h^2=0,223), online games with others (h^2=0,004), writing blogs and texts (h^2=0,076) and the use of skype (h^2=0,114) showed the weakest correlation with all other variables, they were extracted one by one from the list of variables, then the two tests were run again. This time the KMO value (0,825) and the result of the Bartlett's Test (p=0,000), again, justified the adequacy of the variables for the analysis. Results reveal that while using their L1, CLIL learners regarded watching films, texting, learning, family chat, searching for information and watching TV as the most dominant activities in their everyday lives.

Regarding the self-estimated level of L2 skills, the result of the Friedman-test ($X^2(4)=25{,}348$, $p=,000$) reveals no overall statistically significant difference between the mean ranks, that is, participants are on the same opinion. They considered writing (2,56), accent (2,66) and speaking (2,87) as the most, listening comprehension (3,34) and reading (3,57) as the least problematic areas in their second language. The KMO value (0,820) and the Bartlett's Test of Sphericity ($p=0{,}000$) justify the adequacy of variables for the factor analysis. Considering that, the level of self-estimated accent had the lowest communality value ($h^2=0{,}065$) it was extracted from the list of variables, then the two tests were repeated. This time the KMO value (0,819) and the result of the Bartlett's Test ($p=0{,}000$) again, justified the adequacy of the variables for the analysis. The remaining four principal components (speaking, listening comprehension, reading, and writing) are all considered to be present at a remarkable level in the participants' lives.

Regarding the everyday activities in their second language, the result of the Friedman-test ($X^2(12)=351{,}116$, $p=,000$) reveals no overall statistically significant difference between the mean ranks, that is, participants are on the same opinion. Overall, the adequate KMO value (0,574) and the result of the Bartlett's Test ($p=0{,}000$) justify the adequacy of variables for the factor analysis. Considering that, chatting with friends ($h^2=0{,}305$), learning ($h^2=0{,}139$), online games with others ($h^2=0{,}100$), writing blogs and texts ($h^2=0{,}347$), skype ($h^2=0{,}190$), chatting with family ($h^2=0{,}066$) and watching tv ($h^2=0{,}228$) showed the weakest correlation with all other variables, they were extracted one by one from the list of variables, then the two tests were run again. This time the KMO value (0,729) and the result of the Bartlett's Test (Sig. 0,000), again, justified the adequacy of the variables for the analysis. Final results reveal that while using their second language, CLIL learners regarded reading, listening to music, watching films, internet use and info search as the most dominant activities in their everyday lives.

3.1.3. Control group – language presence, language preference

73 respondents provided valid data in terms of their first and second languages. All participants reported Hungarian as their first language (language learnt from birth) and

English as their second language. For some reason 36,98 % (27) of the participants have not answered the question referring to the time spent in the country of their mother tongue or in a second language environment. As the response rate of the control group was low, comparing the two groups based on this question does not yield adequate results, hence only descriptive statistics can be given. According to the valid (46) answers, participants spent about 158 days (nearly 5 months) in the second language environment, while 4709 days (nearly 13 years) in the first language context. While the mean value for the first language covers balanced data, the average number of days spent in the target language environment is made up of extreme values. Many of the participants have not been abroad yet (minimum: 0,00), that is, some outliers (maximum: 2555) increased the mean value. The total numbers (sums) of days spent in the first and second language context reveal a remarkable difference. In case of the first language, it is 29,66 times the number of the days spent in the second language (216615-7302).

Participants' estimation on the presence of the listed languages in their everyday lives revealed a remarkable difference (nearly 43%) between the presence of their first and second languages. Participants reported that they would be more likely to choose their first language ($\bar{x}$ =70,92 percent) over their second language ($\bar{x}$ =27,27 percent). In this case there were no extreme values, 0 percent minimum or 100 percent maximum, for either language, so the data set were compact in each case with balanced standard deviation (SD^{Hun} =14,025, SD^{Eng}=13,035) and variance ($\sigma^{2\ Hun}$=196,707, $\sigma^{2\ Eng}$=169,913).

3.1.4. Control group – L1 and L2 use

The following section reveals control group participants' language preferences in different contexts. In case of reading or listening to something of their interest, Hungarian ($\bar{x}$=64,6301 percent) and English ($\bar{x}$=33,3288) are chosen at different rates, pointing to the fact that the second language is preferred half as much as their mother tongue. In this case there were some extreme values, 0 percent minimum or 100 percent maximum, however standard deviation (SD_{Hun} =27,47752, SD_{Eng}=26,59054) and variance (σ^2_{Hun}=755,014, σ^2_{Eng}=707,057) seem to be balanced.

In case of chatting with someone in the first or the second language the Hungarian language would be chosen at a higher rate ($\bar{x}$=70,8630 percent) than the English language ($\bar{x}$=27,0822 percent). The values of variance (σ^2_{Hun}=536,648, σ^2_{Eng}=473,910), that of standard deviation (SD_{Hun}=23,16566, SD_{Eng}=21,76947) and the sums ($\Sigma_{Hun=}$5173,00, $\Sigma_{Eng=}$1977,00) reveal that some respondents produced extreme values, however it can be finally concluded that control group members seem to consciously choose the first language over the target language for activities like chatting.

In the following sections the results of the applied methods are analyzed and reported in case of the estimated level of language skills and activities being done in the first language. Regarding the self-estimated level of first language skills, the result of the Friedman-test ($X^2(4)$=25701, p=,000) reveals that there is no overall statistically significant difference between the mean ranks, that is, participants are on the same opinion. Control participants considered writing (2,28) as the most and accent (3,29) as the least problematic area in their first language.

The necessary conditions for Principal Components Analysis are met with regard to the adequate KMO value (0,793) and Bartlett's Test of Sphericity (p=0,000) meaning that there is correlation between the variables, hence the model is adequate for the analysis and it should yield reliable factors. Considering that, all the items had an upper value than 0,25, none of them were extracted confirming the strong correlation among all five items. Results reveal that while using their first language, control group learners regarded speaking, listening comprehension, reading, writing and accent as dominant and important skills, among which accent was considered as the least problematic.

Regarding the everyday activities being done in their first language, the result of the Friedman-test ($X^2(12)$=73,971, p=,000) reveals no overall statistically significant difference between the mean ranks. The data show that first language use is the least characterized by writing blogs and texts (3,29), the use of skype (3,45), or playing online games with others (5,50). In contrast, the most frequent activities are internet use (8,62), watching films (8,74) and family chat (9,33)

The condition for Principal Components Analysis, adequate KMO value and significance level are both met. KMO (0,76) and the Bartlett's Test of Sphericity (p=0,002) reveal that the data are adequate for the PCA analysis. Considering that, online games with others (h^2=0,003), writing blogs and text (h^2=0,174), internet use (h^2=0,96) and listening to music (h^2=0,005) showed the weakest correlation with all other variables, they were extracted one by one from the list of variables, then the two tests were run again. This time the KMO value (0,661) and the result of the Bartlett's Test (*p*=0,001), again, justified the adequacy of the variables for the analysis. Results show that while using their first language, control group members regarded watching films, chatting, learning, family chat, skype use, info search, reading and watching TV as the most dominant activities in their everyday lives in their mother tongue.

Regarding the self-estimated level of second language skills, the result of the Friedman-test ($X^2(4)$=13,422 *p*=,009) reveals that there is no overall statistically significant difference between the mean ranks, that is, participants are on the same opinion. They considered speaking (2,75), accent (2,81) and listening comprehension (2,83) as the most, writing (3,13) and reading (3,49) as the least problematic areas in their second language. The KMO value (0,837) and the Bartlett's Test of Sphericity (*p*=0,000) justify the adequacy of variables for the factor analysis. Considering that, the level of self-estimated accent had the lowest communality value (h^2=0,95), referring to the weakest correlation with all other items) it was extracted from the list of variables, then the two tests were run again. This time the KMO value (0,836) and the result of the Bartlett's Test (*p*=0,000) again, justified the adequacy of the variables for the analysis. The remaining four principal components (speaking, listening comprehension, reading and writing) are all considered to be present at a remarkable level in the participants' lives.

Regarding the activities being done in their second language, the result of the Friedman-test ($X^2(12)$=376,262 *p*=,000) reveals that there is no overall statistically significant difference between the mean ranks, that is, participants are on the same opinion. Overall, the adequate KMO value (0,733) and the result of the Bartlett's Test (*p*=0,000) justify the adequacy of variables for the factor analysis.

Considering that, watching TV (h^2=0,233), chatting with family (h^2=0,161), learning (h^2=0,055), listening to music (h^2=0,210) showed the weakest correlation with all other variables, they were extracted one by one from the list of variables, then the two tests were run again. This time the KMO value (0,794) and the result of the Bartlett's Test (*p*=0,000), again, justified the adequacy of the variables for the analysis. Final results reveal that while using their second language, control group members regarded reading, watching films, internet use, chatting, online games, writing texts and blogs, skype use and info search as the most dominant activities in their everyday lives. Table 11 shows CLIL and control group participants' self-assessed strengths, weaknesses, and dominant activities in their mother tongue and the second language. Table 9 and 10 show data of language presence, preference and self-assessed skills in case of Hungarian (as an L1) and English (as an L2).

	CLIL	CONTROL
First language	Hungarian	Hungarian
Second language	English	English
Days (24 hours) spent in the first language context	4550	4709
Days (24 hours) spent in the second language context	224	158
Difference of language use between L1 and L2	18.2% (57.86-39.66)	43.65% (70.92-27.27)
Difference of listening-reading activities between L1 and L2	5.95% (51.45-45.5)	31.31% (64.63-33.32)
Difference of talking in L1 and L2	17.76% (57.76-40)	43.78% (70.86-27.08)

Table 9: Self-assessed language parameters

	Speaking (%)		Listening (%)		Reading (%)		Writing (%)	
	L1	L2	L1	L2	L1	L2	L1	L2
CLIL	80.55	71.90	84.72	76.36	83.36	77.72	77.55	70.27
Control	86.63	59.18	87.18	59.90	85.72	65.81	80.82	62.54

Table 10: Self-assessed level of L1 and L2 skills

Table 11 reveals the self-assessed strengths and weaknesses in the L1 and the L2 along with the different activities they are used for:

		Order of activities: from the most... to the least ...	**CLIL group**	**Control group**
HUNGARIAN	**SKILLS**	Strengths	listening, (accent)	accent, (listening)
		Weaknesses	writing, speaking, reading	writing, reading, speaking
		The most dominant	listening, writing, speaking, (reading)	reading, speaking, listening, writing, (accent)
	ACTIVITIES	Frequent	family chat, learning, (chatting)	family chat, watching films, (internet use)
		Rare	online games, writing texts, (Skype use)	online games, Skype use, (writing texts)
		The most dominant	chatting, family chat, watching TV, watching films, info search, (learning)	watching films, learning, info search, reading, family chat, watching TV, (chatting)
ENGLISH	**SKILLS**	Strengths	reading, listening	reading, writing
		Weaknesses	writing, accent, speaking	speaking, accent, listening
		The most dominant	writing, listening, speaking, (reading)	writing, listening, speaking, (reading)
	ACTIVITIES	Frequent	listening to music, internet use, (info search)	listening to music, internet use, (online games)
		Rare	family chat, Skype use, (writing texts)	writing texts, Skype use, (family chat)
		The most dominant	internet use, info search, listening to music, reading, (watching films)	internet use, chatting, info search, reading, writing texts, online games, Skype use, (watching films)

Table 11: Summary table on L1 and L2 use

3.2. Test d2-R – test of selective attention

Data was processed by SPSS software. 142 learners produced evaluable data of whom 69 were CLIL learners and 73 were controls. There were no incomplete, unidentifiable or invaluable test sheets. Data were recorded block by block then summed for further analysis. All the 798 signs of the d2-R test are arranged in 4 blocks (three rows constitute a block that is repeated four times). Figure 11 shows how participants' mean performance values altered during the four blocks.

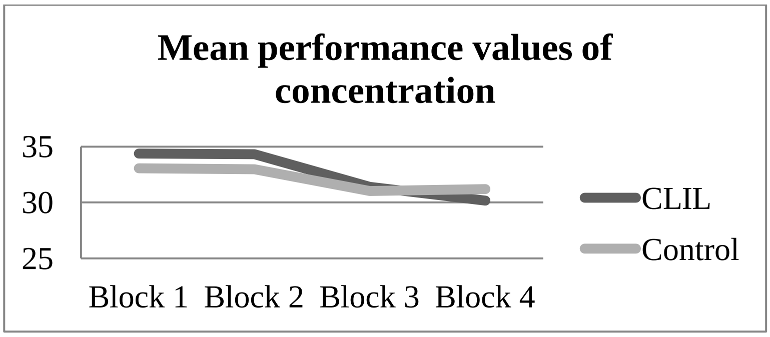


Figure 11: Mean performance values of concentration

In the first block CLIL learners processed *M*=38 target stimuli and made *M*=3,36 omission errors and *M*=0,26 incorrect markings. In the second block they processed *M*=38,34 target stimuli and made *M*=3,75 omission errors and *M*=0,27 incorrect markings. In the third block they processed *M*=34,85 target stimuli and made *M*=3,23 omission errors and *M*=0,23 incorrect markings. Finally, in the fourth block they processed *M*=35,27 target stimuli and made *M*=4,56 omission errors and *M*=0,52 incorrect markings.

The control group participants processed *M*=38,12 target stimuli and made *M*=4,43 omission errors and *M*=0,63 incorrect markings in the first block. In the second block they processed *M*=38,41 target stimuli and made *M*=4,8 omission errors and *M*=0,61 incorrect markings. In the third block they processed *M*=36,13 target stimuli and made *M*=4,54 omission errors and *M*=0,54 incorrect markings. Finally, in the fourth block they processed *M*=36,68 target stimuli and made *M*=4,8 omission errors and *M*=0,64 incorrect markings.

The summary table 12 shows the mean number of processed target stimuli, omission errors and incorrect markings made by the two groups during the four blocks:

	CLIL (Mean)				Control (Mean)			
	Block 1	Block 2	Block 3	Block 4	Block 1	Block 2	Block 3	Block 4
Processed target stimuli	38	38.34	34.85	35.27	38.12	38.41	36.13	36.68
Omission errors	3.36	3.75	3.23	4.56	4.43	4.8	4.54	4.8
Incorrect markings	0.26	0.27	0.23	0.52	0.63	0.61	0.54	0.64

Table 12: Mean d2-R results

The difference between the total number of errors (omission errors and incorrect markings together) made by the two groups might seem remarkable for the first sight. The total number of (both) error types was Σ=1138 in case of the CLIL group, and that of the control group it was Σ=1536. Figure 12 shows the total number of errors made by the two groups during the four blocks:

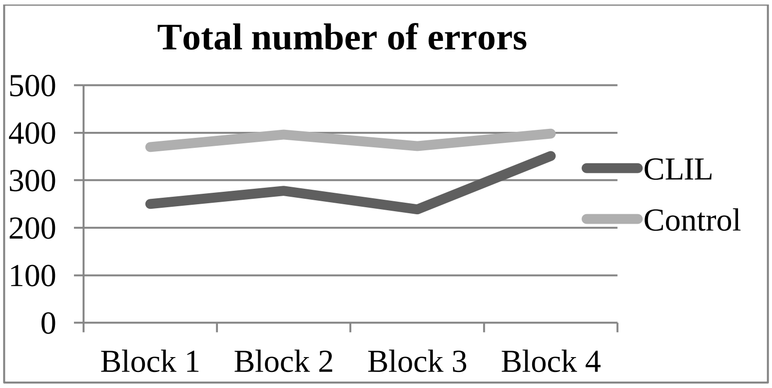


Figure 12: Total number of errors

22,36 percent of the total errors was made by the CLIL participants in the first block, 24,86 percent in the second, 21,37 percent in the third and 31,39 in the fourth. For the first sight these results might seem to be hectic in comparison to those of the control group. In case of the controls 24,06 percent of the total errors was made in the first block, 25,78 in the second, 24,21 in the third and 25,91 in the fourth.

Since the related graphs are skewed to right they reveal that nearly half of the groups made no errors during the task and there were only a few participants who produced unusual values (outliers).

To reveal whether there is a significant difference between the two groups the F-test and the Levene-test were applied. The significance level of the probability for the F-test (that is sensitive to distribution) is 0,166 (p=,166). The result of the Levene's test (for the homogenity of variances) is in line with that of the F-test (p=,199). Both results are above 0,05, consequently, the null hypothesis is accepted, that is the condition of homogenity of variances has been satisfied, meaning that there is no significant difference between the standard deviations of the variables.

In the next step the distribution of further data is investigated. Since no normal distribution in case of the partial or final (total) results is observed, the focus of attention is shifted to the further investigation of the direction and extent of the deviation from normal distribution.

Although data in the summary table (see Annex ...) seem to be close to one another, the overall picture shows otherwise. Examination of skewness makes distorting data effects (causing non-normally distributed bell curves) visible on histograms. Generally, it can be stated that extreme variability in the data was not typical, although in both groups there were some participants (a core group) whose general performance was nearly close to the average. In case of the CLIL participants the test for skewness (with regard to all blocks and for the total) resulted in a distribution skewed to right, meaning that the number of lower values is frequent, while that of high values is rare. The number of outliers in case of this group has a more remarkable effect on the total performance. However, the direction of data bias here is positive and compensates for the lower performance of a larger core group.

In contrast, control group results showed a different picture in the test for skewness. In this case the distribution is skewed to left indicating that the number of lower values is rare, while that of high values is frequent. Here it can be concluded that some learners (a core group) produced lower values than the average in the second, third and fourth blocks and none of the learners were able to compensate for these results. The histograms below show the performance values of the control group in the four blocks and also the total performance values for concentration.

Considering that the lack of normal distribution precludes parametric statistical tests, only non-parametric tests could have been implemented. To decide whether there is a significant difference between the performance values of concentration between the two groups, F-test was applied (as the only accepted test type in Sajtos-Mitev (2007) in case of non-normal distributions and skewness). F-test is used for defining the exact difference between the variances of the samples. In this particular case, the variances of the two groups are the following: $\sigma^2_{CLIL}=25{,}174^2=633{,}732$ és a $\sigma^2_{control}=24{,}847^2=617{,}389$. If F value is close to 1, H_0 (there is no difference between the variances of the two groups) is accepted. In this case F=1,026 (p=0,353) so it can be concluded that there is no significant difference between the performance values of concentration of the two groups.

Finally, learners' performance was compared to that of their age group. Consequently, the mean performance values of attention (concentration) of the CLIL group is average: (CLIL=130,27; 100 standard points). With regard to the accuracy of the measurement device (confidence intervals 95,6-104,4, p=0,05), CLIL learners' performance value of concentration is average. Their work pace (number of processed target stimuli; 100 standard points) is also average even taking into account the measurement error of the test (confidence intervals: 95,2-104,8, p=0,05). During task solution they worked with average accuracy (H%=98 standard points). With regard to the measurement error of the test (confidence intervals: 90,7-105,3, p=0,05) their accuracy is low to average.

The mean performance values of attention (concentration) of the control group is average: (control=128,4; 99 standard points). With regard to the accuracy of the measurement device (confidence intervals 94,6-103,4, p=0,05), control group learners' performance value of concentration is low to average. Their work pace (number of processed target stimuli; 101 standard points) is also average even taking into account the measurement error of the test (confidence intervals: 95,2-105,8, p=0,05). During task solution they worked with average accuracy (H%=95 standard points). With regard to the measurement error of the test (confidence intervals: 87,7-102,3, p=0,05) their accuracy is low to average. The data in Table 13 are based on the d2-R Test Manual (Brickenkamp et al., 2010)

	CLIL group			Control group		
	H%[1]	PTS[2]	Attention	H%	PTS	Attention
Mean	11,05	146,47	130,27	14,01	149,38	128,4
Standard points	98	100	100	95	101	99
Confidence intervals	90,7-105,3	95,2-104,8	95,6-104,4	87,7-102,3	96,2-105,8	94,6-103,4
Evaluation	(low to) average	average	average	(low to) average	average	(low to) average

Table 13: Comparison of d2-R test outcomes to the normative data

3.3. Phonemic fluency performance

In the following sections we investigate whether there is a significant difference between the total number of words generated by the two groups for both test conditions (FAS and KTA).

3.3.1. Quantity of words – results within groups

The two groups produced 3007 words in total in the English verbal fluency test (F, A, S) and 4118 words in the Hungarian verbal fluency test (K, T, A). Table 14 reports on the distribution of generated words between the groups:

	FAS	KTA	Total
CLIL group	1681	1917	3598
Control group	1326	2201	3527
Total	3007	4118	7125

Table 14: Total number of generated words in the fluency tests

[1] H% (accuracy)

[2] PTS (processed target stimuli)

3.3.2. CLIL group results

CLIL participants generated 1,681 words of the total in the English language test of which 592 words begin with F, 442 words with A and 647 words with S. Regarding the Hungarian language they produced 1917 words in total of which 685 words begin with K, 663 words with T and 569 words with A. The ratio of words written in Hungarian in comparison to words written in English is 1.14. Tests of normality (Kolmogorov-Smirnov and Saphiro-Wilk tests) revealed normal distribution only in case of letter T in the Hungarian test (p=.075 and p=.051) hence the null hypothesis is accepted. In case of the remaining letters (F: p=.000, p=.003; A: p=.011, p=.000; S: p=.000, p=.000); K: p=.053, p=.028; A: p=.014; p=.011) the null hypothesis must be rejected (p<.05). Given that the probability of randomness in these test results is lower than 0.05 regarding the Kolmogorov-Smirnov and Saphiro-Wilk tests (p<.05) it can be concluded that the ratio of words in Hungarian and English (1.14) cannot be attributed to randomness either, hence it is statistically significant. Table 15 presents the total number of words produced by the CLIL group in the two tests:

CLIL group	English phonemic fluency test (FAS)						Hungarian phonemic fluency test (KTA)					
	F		A		S		K		T		A	
Tests of Normality	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk
	.000	.003	.011	.000	.000	.000	.053	.028	.075	.051	.014	.011
Number of words	592		442		647		685		663		569	
TOTAL	1681						1917					

Table 15: Normality test results on the total number of words generated by the CLIL group in the fluency tests

3.3.3. Control group results

Control participants produced 1326 words of the total of which 513 words begin with F, 283 words with A and 530 words with S. Regarding the Hungarian language they produced 2201 words in total of which 765 begin with K, 782 with T and 654 with A.

The ratio of words written in Hungarian in comparison to words written in English is 1.66. Tests of normality (Kolmogorov-Smirnov and Saphiro-Wilk tests) revealed normal distribution in case of letter S in the English test (p=.200 and p=.453) and letter A in the Hungarian test (p=.095 and p=.092), that is the null hypothesis in these cases is accepted. In case of the remaining letters (F: p=.024, p=.324; A: p=.000, p=.001; K: p=.000, p=.016; T: p=.010, p=.181) the null hypothesis is rejected ($p<.05$). Given that the probability of randomness in these test results is lower than 0.05 (in accordance with the indicators of significance in the Kolmogorov-Smirnov and Saphiro-Wilk tests) it can be concluded that the ratio of words in Hungarian and English (1.66) cannot be attributed to randomness, either, hence it is statistically significant. Table 16 presents the total number of words produced by the control group in the two tests:

Control group	English phonemic fluency test (FAS)						Hungarian phonemic fluency test (KTA)					
	F		A		S		K		T		A	
Number of words	513		283		530		765		782		654	
Tests of Normality	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk
	.024	.324	.000	.001	.200	.453	.000	.016	.010	.181	.095	.092
TOTAL	1326						2201					

Table 16: Normality test results on the total number of words generated by the control group in the fluency tests

3.3.4. Quantity of words - results between groups

The total number of words generated in the English phonemic fluency tests by the CLIL group is1681, and 1326 in case of the control group. The CLIL group participants generated 1.27 times as many words as the control group participants. To reveal whether this number is significantly different Mann-Whitney test was run. It can be concluded from the results of the Mann-Whitney test that there is a significant difference ($p<.05$) between the two groups for all three test conditions (F: p=.004; A: p =.000; S: p =.000)

The number of total words generated in the Hungarian phonemic fluency test by the CLIL group is 1917, and 2201 in case of the control group. The control group participants generated 1.15 times as many words as the control group participants. Due to the fact that the data in the three test conditions (K, T, A) are not normally distributed in either group, the non-parametric Mann-Whitney test was applied. In contrast to the results for the English phonemic fluency test, this time the outcomes are more varied. The p-value of .042 for letter T indicates a significant difference between the medians of the two groups. However, in case of letter K (p=.591) and A (p=.168) the p-value is higher than 0.05, therefore no significant group differences can be detected. Table 17 reports on the results of the Mann-Whithey tests for each letters:

	Mann-Whitney
F	***.004***
A	***.000***
S	***.000***
K	.591
T	***.042***
A	.168

Table 17: Group differences in the total words

To reveal whether there is a significant difference between the total number of words for the Hungarian test conditions (KTA) a deeper analysis has been done. Tests of normality (Kolmogorov-Smirnov, Shpiro-Wilk) showed significant results, $p<.05$, that is no normal distribution can be observed (CLIL group: p=.092; p=.004; control group (p=.007, p=.021). Since the Levene's test for the equality of variances showed significant differences between the variances of the two groups (p=.018), thus the non-parametric Mann-Whitney Test could have been applied. Table 18 reveals the results of the Mann-Whitney Test, that is, no significant difference between the two groups in terms of the total number of words created in the Hungarian phonemic fluency test (p=.200) has been found as it is seen in Table 18:

		CLIL	Control
Tests of Normality	Kolgomorov-Smirnov	.092	.007
	Shapiro-Wilk	.004	.021
Levene's test	.018		
Mann-Whitney	.200		

Table 18: Group differences in the Hungarian fluency test (KTA)

3.3.5. Clusters

Four types of clusters (phonological slight, phonological strict, semantic slight and semantic strict) have been differentiated in accordance with the corresponding scientific literature (Troyer, et al., 1997; Mészáros, et al., 2011). In this section we aim to reveal whether there is a significant difference between the number of different cluster types created by the groups and the mean cluster sizes.

First, the total number of clusters created by the two groups in the two tests (FAS+KTA) is investigated. Table 19 shows the baseline statistics for the total number of clusters created for each evaluation criteria (cluster types):

FAS+KTA (English test + Hungarian test)	Phonological cluster		Semantic cluster		Mean cluster size
	Slight	Strict	Slight	Strict	
CLIL	412	149	218	81	12.97
Difference	1.02x↑	1.15x↑	1.09↓	1.37↑	1.10↑
Control	403	129	238	59	11.72

Table 19: Total number of clusters created by the two groups in all tests

Baseline statistics show the dominance of the CLIL group in case of strict phonological and semantic evaluation criteria (clusters), and that of the control group in case of slight semantic evaluation criterion (cluster) for the Hungarian and the English fluency tests.

3.3.5.1. Number of clusters – English phonemic fluency tests (FAS)

In the following sections the differences between the numbers of clusters created by both groups in the English phonemic fluency tests are investigated. Table 20 .shows the total number of clusters generated by both groups in the English phonemic fluency test and the mean cluster sizes:

FAS (English test)	Phonological clusters		Semantic clusters		Mean cluster size
	Slight	Strict	Slight	Strict	
CLIL	241	80	51	37	6.08
Difference	1.27x↑	2.10x↑	1.02x↓	4.11x↑	1.27x↑
Control	189	38	52	9	4.76

Table 20: Total number of clusters generated by both groups in the English fluency tests

The arrows in the Table 20 indicate the direction of higher productivity. For the first sight a remarkable difference can be detected under the strict evaluation criteria for both cluster types. CLIL group participants created 2.10 times as many strict phonological clusters and 4.11 times as many strict semantic clusters as the control group members. They also generated 1.27 times as many slight clusters than the members of the control group. In the next step the distribution of the data related to the separate letters are investigated. Both tests of normality (Kolmogorov-Smirnov and Saphiro-Wilk tests) for all cluster types showed *p*=.000, that is lower than 0.05 assuming non-normal distributions. The same non-normal distribution has been found in case of the mean cluster size, also: *p*=.000. Therefore, to reveal whether there is a statistically significant difference between the two groups with regard to the number of generated clusters for all letter conditions (F, A, S), the non-parametric Mann-Whitney test was applied. Table 21 shows the results (p-values) of the Mann-Whitney tests in case of cluster types and the mean cluster size in the English fluency tests:

FAS (English test)	Phonological clusters		Semantic clusters		Mean cluster size
	Slight	Strict	Slight	Strict	
F	.272	.052	.899	**.002**	**.006**
A	**.003**	.130	.210	.071	**.001**
S	.125	**.005**	.739	**.030**	**.008**

Table 21: Group differences on clusters in the English fluency tests

The highlighted p-values in Table 21 indicate significant differences between the number of clusters created by the CLIL group and the control group participants. The CLIL group participants produced significantly more slight phonological clusters in case of letter A. They also produced significantly more strict phonological clusters in case of letter S. Considering the number of strict semantic clusters CLIL participants created more of them in case of letter F and S. CLIL group participants have significant differences in case of mean cluster size for each letter in the English phonemic fluency test.

3.3.5.2. Number of clusters – Hungarian phonemic fluency tests (KTA)

In the following sections the differences between the numbers of clusters are investigated. Table 22 shows the total number of clusters generated by both groups in the Hungarian phonemic fluency tests:

KTA (Hungarian test)	Phonological clusters		Semantic clusters		Mean cluster size
	Slight	Strict	Slight	Strict	
CLIL	171	69	167	44	6.88
Difference	1.25x↓	1.32x↓	1.11x↓	1.14x↓	1.03x↓
Control	214	91	186	50	7.1

Table 22: Total number of clusters generated by both groups in the Hungarian fluency tests (KTA)

The arrows in the Table 22 indicate the direction of higher productivity in favour of the control group for all evaluation criteria (clusters). The control group participants created 1.25 times as many slight phonological clusters than the participants of the CLIL group. They also created 1.32 times more strict phonological, 1.11 times more slight semantic clusters and 1.14 times more strict semantic clusters than the experimental group. Tests of normality (Kolmogorov-Smirnov and Saphiro-Wilk tests) revealed non-normal distribution in any of the clusters types (p=.000). The same non-normal distribution has been found in case of the mean cluster size, also (p=.000). Since parametric tests can only be applied in case of normal distribution, this time the non-parametric Mann-Whitney test was applied again, to reveal whether there is a statistically significant difference between the two groups with regard to the number of generated clusters for all letter conditions (K, T, A). Table 23 shows the results (p-values) of the Mann-Whitney tests regarding the different letters, cluster types and the mean cluster size:

KTA (Hungarian test)	Phonological clusters		Semantic clusters		Mean cluster size
	Slight	Strict	Slight	Strict	
K	**.007**	.709	.448	.070	.661
T	.888	.172	.266	.069	.065
A	.554	.741	.706	.738	.680

Table 23: Group differences on clusters in the Hungarian fluency tests (KTA)

The highlighted p-value in Table 23 indicates a significant difference between the number of slight phonological clusters created by the groups. The control group created significantly more slight phonological clusters in case of letter K in the Hungarian phonemic test compared to the CLIL group. In terms of the other cluster types no significant difference between the groups could have been detected.

3.3.6. Switches

In accordance with the corresponding literature (Abwender, 2001) two types of switches (hard switches and cluster switches) have been differentiated in the theoretical part of the dissertation. In the following sections we aim to reveal whether there is a significant difference between the numbers of switches in case of both groups. First, the baseline statistics are reported in Table 24:

	CLIL group		Control group	
	Cluster switches	Hard switches	Cluster switches	Hard switches
F	53	295	33	292
A	29	247	5	175
S	70	321	38	312
FAS Σ	**152**	**863**	**76**	**779**
K	70	330	69	368
T	55	322	76	349
A	52	289	38	359
KTA Σ	**177**	**941**	**183**	**1076**
TOTAL	**329**	**1804**	**259**	**1855**

Table 24: Baseline statistics on the different types of switches

The total number of clusters created by the CLIL group (329) is 1.27 times as many as that of the control group (259). This dominance is caused by the number of cluster switches produced by the CLIL group for each letter (F, A, S). In contrast, the number of hard switches seems to be dominant in case of the control group: the difference between the two sums is approximately 3% (2.749). This difference might be explained by the number of hard switches made by the control group for each letter (K, T, A). In the next sections we aim to reveal whether these data indicate significant differences.

First, the distribution of data (total number of cluster switches and hard switches) in case of the CLIL group is investigated. Tests of normality (Kolmogorov-Smirnov and Saphiro-Wilk tests) revealed significant p-values (lower than 0.05; *p*=.000 and *p*=.000) that indicate non-normal distribution for cluster switches in case of CLIL learners. Regarding hard switches the normality tests revealed no significant p-values (*p*=.200 and *p*=.077), that indicate normal distribution for hard switches in case of CLIL learners.

With regard to the control group, tests of normality revealed significant p-values, that indicate non-normal distribution for either switch type (cluster switches: *p*=.000; *p* =.000; hard switches: *p* =.004, *p* =.022). For this reason, the non-parametric Mann-Whitney test has been run to reveal whether there is a significant difference between the two groups in terms of cluster switches and hard switches. Table 25 shows the results (p-values) of the Mann-Whitney tests regarding the different switches in the English and Hungarian test types.

	Cluster switches	Hard switches
F	.133	.616
A	***.001***	***.000***
S	***.022***	.326
K	.909	.700
T	.179	.918
A	.242	.063

Table 25: Group differences in cluster switches and hard switches

The highlighted p-values in Table 25 indicate significant differences between the numbers of switches made by the groups.

CLIL group participants made significantly more cluster switches (p=.001) and hard switches (p=.00) in letter A (in the English phonemic fluency test) and more cluster switches in case of letter S (p=.022) also in the English phonemic fluency test. Although, baseline statistics suggested the control group's dominance in the number of hard switches for the Hungarian test (KTA) version, it is not statistically confirmed.

Finally, we calculated the difference between the two groups in terms of the total number of switches produced by the groups in the two tests (Hungarian: KTA and English FAS). Table 26 reports on the results of the Normality tests in case of the CLIL group. Significant p-values (p<.05) refer to non-normal distributions. Table 27 reports on the results of the Normality tests in case of the Control group. Significant p-values (p<.05) refer to non-normal distributions:

CLIL	**Cluster swites**		**Hard switches**	
	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk
FAS	***.000***	***.000***	.051	.083
KTA	***.003***	***.004***	.200	.422

Table 26: Normality test results on switches in both fluency tests (CLIL)

Control	**Cluster switches**		**Hard switches**	
	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk
FAS	***.000***	***.000***	***.035***	.195
KTA	***.000***	***.000***	.200	.326

Table 27: Normality test results on switches in both fluency tests (control)

Since in addition to normal distribution, homogeneity of variances is also a necessary condition for the independent samples T-test it has also been calculated. Its results are: cluster switches (FAS): F (1; 140) p=.001; cluster switches (KTA): F (df2:140) p=.222; hard switches (FAS): F (1; 140) p=.485; hard switches (KTA): F (1; 140) p=.460. If p<0.05 than the null hypothesis is accepted, meaning that there is significant difference between the variances of the two groups. For this reason, the independent samples T-test can be applied in case of hard switches for the Hungarian (KTA) and English (FAS) tests.

In case of cluster switches (for both language tests) the non-parametric Mann-Whitney test has been applied. The independent-samples t-test results are as follows: significant difference in the number of hard switches has been found for the FAS test condition: $t(140)=2.427$, $p=.016$, but not for the KTA test condition: $t(140)=-1.408$, $p=.161$. These results suggest that there is a significant difference between the performances of the two groups in terms of the hard switches for the English phonemic fluency test. Regarding cluster switches for both tests (FAS and KTA) the Mann-Whitney test has been applied. It can be concluded from the test that there is a significant difference ($p<.05$) between the performances of the two groups in terms of cluster switches for the FAS test ($p=.000$) while no significant difference can be observed between the groups in terms of cluster switches for the KTA test. The highlighted p-values in Table 28 indicate significant differences between the two groups in terms of cluster switches and hard switches:

	Cluster switches (Mann-Whitney)	Hard switches (Independent-Samples T test)
FAS	***.000***	***.016***
KTA	.541	.161

Table 28: Group differences on the different switches

We can conclude that CLIL group participants created significantly more cluster switches and hard switches in the English (FAS) fluency test, however, there is no significant difference between the two groups in terms of these switches in the Hungarian (KTA) fluency test.

3.3.7. Errors, repetitions, and perseverations

In the following sections we aim to reveal the differences between the CLIL and the control group in terms of errors, repetitions, and perseverations. Errors are defined as words that begin with a wrong letter, are proper nouns, geographical names or have suffixes. Perseverations are defined as the immediate appearance of the same word, while repetitions denote a later reappearance of a particular word (Troyer, 2000). First, baseline statistics to the related parameters are reported in Table 29:

	CLIL group			Control group		
	Errors	Repetitions	Perseverations	Errors	Repetitions	Perseverations
F	33	0	1	24	1	1
A	13	0	0	17	3	0
S	25	1	0	24	0	0
FASΣ	**71**	**1**	**1**	**65**	**4**	**1**
K	4	3	0	6	0	0
T	6	1	0	4	1	0
A	9	0	0	11	1	0
KTAΣ	**19**	**4**	**0**	**21**	**2**	**0**
TOTAL	**90**	**5**	**1**	**86**	**6**	**1**

Table 29: Baseline statistics on errors, repetitions and perseverations

Table 29 reports that CLIL group participants made more errors in comparison to the control group participants, however, it is important to note that their productivity (fluency) in the English phonemic fluency test was higher also (CLIL:1681; control group:1326). Contrarily, the number of errors made by the control group participants in the Hungarian phonemic fluency test seems to be higher. In terms of repetitions there is only a slight difference between the two groups. Regarding perseverations the data are extremely low and completely identical for the two groups.

First, the distribution of data (total number of errors, repetitions and perseverations) are investigated separately in case of the CLIL group. Tests of normality (Kolmogorov-Smirnov and Saphiro-Wilk tests) revealed significant p-values (lower than 0.05) that indicate non-normal distribution for the number of errors (p=.000 and p=.000), repetitions (p=.000 and p=.000) and perseverations (p=.000 and p=.000) in case of CLIL learners. The same results have been found (p=.000 and p=.000) in case of the control group for all three parameters. For this reason, the non-parametric Mann-Whitney test has been run to reveal whether there is a significant difference between the two groups in terms of errors, repetitions and perseverations. Table 30 shows the results (p-values) of the Mann-Whitney tests regarding the three parameters in the English and Hungarian test types:

	Errors	Repetitions	Perseverations
F	.310	.331	.968
A	.321	.331	1.000
S	.917	.304	1.000
K	.783	.144	1.000
T	.225	.968	1.000
A	.693	.331	1.000

Table 30: Group differences on errors, repetitions and perseverations

The highlighted p-values in Table. 30 would indicate statistically significant differences. Consequently, it can be concluded that there is no statistically significant difference between the CLIL group and the control group with regard to the number of errors, repetitions and perseverations made in the English and Hungarian phonemic fluency tests. These results are in line with our expectations.

3.3.8. Word classes

First, the total number of word classes created by the two groups in the two tests (FAS+KTA) is investigated. In Table 31 below baseline statistics report the dominance of the control group in terms of the total number of concrete nouns, verbs and other word types produced in both test version (Hungarian and English phonemic fluency tests). Conversely, the CLIL group participants generated 1.70 times as many abstract nouns and 1.31 times as many adjectives as the other group. Table 31 summarizes the total number or generated words with regard to word classes in both tests.

FAS+KTA (English test + Hungarian test)	Abstract nouns	Concrete nouns	Verbs	Adjectives	Other word types
CLIL	309	2112	508	388	281
	1.70x↑	1.04x↓	1.05x↓	1.31x↑	1.10x↓
Control	181	2205	536	295	310

Table 31: Total number of generated words in the word classes in both fluency tests

In the next sections, following the logic so far, the focus is on the individual examination of the languages. In case of the English phonemic fluency test CLIL group dominance can be detected in all but one word types. These data can be seen in Table 32:

FAS (English phonemic fluency test results)					
	Abstract nouns	Concrete nouns	Verbs	Adjectives	Other word types
CLIL group	145	814	306	255	161
Difference	3.22x↑	1.31x↑	1.06x↑	1.53x↑	1.32x↓
Control g.	45	620	284	164	213

Table 32: Total number of generated words in the word classes in the English fluency test

To reveal whether there is a statistically significant difference between the two groups in terms of all word types, first the distribution of the data related to the different letters are examined. Tests of normality (Kolmogorov-Smirnov and Shapiro-Wilk tests) revealed the following data: F (*p*=.000, *p*=.000), A (*p*=.000, *p*=.000) and S (*p*=.000, *p*=.000) in case of the word types. Regarding CLIL group and F(*p*=.000, *p*=.000), A (*p*=.000, *p*=.000) and S (*p*=.000, *p*=.000) in case of the control group for all word types. Since all the results showed significant differences ($p<0.05$), it can be concluded the data are not normally distributed. For this reason, the non-parametric Mann-Whitney test was applied. Table 33 shows the results (p-values).

FAS (English test)	Abstract nouns	Concrete nouns	Verbs	Adjectives	Other word types
F	**.008**	**.040**	.574	**.001**	**.011**
A	**.000**	**.000**	**.003**	**.004**	.484
S	**.000**	**.001**	.724	.061	.319

Table 33: Group differences on word classes

The highlighted p-values in Table 33 indicate significant differences between the numbers of different word types generated by the two groups.

CLIL group members generated significantly more abstract nouns (F, A, S), concrete nouns (F, A, S), verbs (A), and adjectives (F, A). However, the control group produced significantly more words that belong to other word types (F).

Regarding the Hungarian phonemic fluency test baseline statistics assume the dominance of the control group in terms of the total number of concrete nouns and that of verbs. Conversely, the CLIL group participants generated 1.64 times as many abstract nouns, 1.01 times as many adjectives and 1.23 times as many other word types as the other group. These data is reported in Table 34:

KTA (Hungarian phonemic fluency test results)					
	Abstract nouns	Concrete nouns	Verbs	Adjectives	Other word types
CLIL group	164	1298	202	133	120
Difference	1.20x↑	1.22x↓	1.24x↓	1.01x↑	1.23x↑
Control g.	136	1585	252	131	97

Table 34: Total number of generated words in the word classes in the Hungarian fluency tests

To reveal whether these results are statistically significant, first the distributions of the data related to the different letters are examined. Table 35 reports on the results of the tests of normality (Kolmogorov-Smirnov and Shapiro-Wilk tests) in case of the CLIL group. In case of significant differences ($p<.05$), meaning that non-normal distributions are found.

	K		**T**		**A**	
CLIL	Kolmogorov-Smirnov	Shapiro-Wilk	Kolmogorov-Smirnov	Saphiro-Wilk	Kolmogorov-Smirnov	Shapiro-Wilk
abstract nouns	***.000***	***.000***	***.000***	***.000***	***.000***	***.000***
concrete nouns	***.042***	.344	***.023***	.074	***.046***	.197
verbs	***.000***	***.000***	***.000***	***.000***	***.000***	***.000***
adjectives	***.000***	***.000***	***.000***	***.000***	***.000***	***.000***
other word types	***.000***	***.000***	***.000***	***.000***	***.000***	***.000***

Table 35: CLIL group's normality test results for word classes in the Hungarian fluency tests

Table 36 reports on the results of the tests of normality (Kolmogorov-Smirnov and Saphiro-Wilk tests) in case of the Control group. In case of significant differences (p<.05), indicating non-normal distributions.

	K		**T**		**A**	
Control	Kolmogorov-Smirnov	Saphiro-Wilk	Kolmogorov-Smirnov	Saphiro-Wilk	Kolmogorov-Smirnov	Saphiro-Wilk
abstract nouns	***.000***	***.000***	***.000***	***.000***	***.000***	***.000***
concrete nouns	***.010***	***.010***	.200	.074	***.000***	***.017***
verbs	***.000***	***.000***	***.000***	***.000***	***.000***	***.000***
adjectives	***.000***	***.000***	***.000***	***.000***	***.000***	***.000***
other word types	***.000***	***.000***	***.000***	***.000***	***.000***	***.000***

Table 36: Control group's normality test results for word classes in the Hungarian fluency tests

In the next step the non-parametric Mann-Whitney test was applied to compare the data of the two groups. Table 37 contains the results of the test:

KTA (Hungarian test)	Abstract nouns	Concrete nouns	Verbs	Adjectives	Other word types
K	.621	.111	.205	.741	.233
T	.322	**.029**	.652	.912	.660
A	.231	**.021**	**.004**	.925	.169

Table 37: Group differences in terms of word classes in the Hungarian fluency tests

The highlighted p-values in Table 37 indicate significant differences between the numbers of different word types generated by the two groups. That is, the control group generated significantly more concrete nouns and verbs beginning with letter A and concrete nouns beginning with letter T. Regarding the remaining letters and word types no significant difference has been found between the groups

Table 38 summarizes those parameters under which significant differences between the groups have been found. 'CLIL' and 'CON' indicate the significant performance of the corresponding group, while '-' relates to no statistically significant differences between the groups.

		F	A	S		K	T	A
ALL WORDS		CLIL	CLIL	CLIL		CON	-	-
CLUSTERS	Phonological Slight	-	CLIL	-		CON	-	-
	Phonological Strict	-	-	CLIL		-	-	-
	Semantic Slight	-	-	-		-	-	-
	Semantic Strict	CLIL	-	CLIL		-	-	-
MEAN CLUSTER SIZE		CLIL	CLIL	CLIL		-	-	-
SWITCHES	Cluster	-	CLIL	CLIL		-	-	-
	Hard	-	CLIL	-		-	-	-
'ERRING'	Errors	-	-	-		-	-	-
	Repetitions	-	-	-		-	-	-
	Perseverations	-	-	-		-	-	-
WORD TYPES	Abstract nouns	CLIL	CLIL	CLIL		-	-	-
	Concrete nouns	CLIL	CLIL	CLIL		-	CON	CON
	Verbs	-	CLIL	-		-	-	CON
	Adjectives	CLIL	-	CLIL		-	-	-
	Other	CON	-	-		-	-	-

Table 38: Summary table on significant findings related to fluency tests

Finally, some of our results on the different variables of the Hungarian verbal fluency test could be compared to the normative data of Hungarian monolingual children of the same age group tested by Tánczos (2014). Since in this study, our scoring principle regarding the minimum number of words in a cluster was not in line with Tánczos's (2014), comparing these data based on a different scoring principle would not be grounded statistically. The comparable variables are seen in Table 39:

Variables	Letters	CLIL	Control	Normative data (Tánczos; 2014)
Individual performance	K	9.92	10.48	11.83
	T	9.59	10.74	10.70
	A	8.26	8.93	8.74
	F	8.57	7.02	-
	A	6.40	3.87	-
	S	9.37	7.26	-
Total number of words	K	685	765	611.82
	T	662	784	577.8
	A	570	652	471.96
	F	592	513	-
	A	442	283	-
	S	647	530	-
Error indicator	K	0.005	0.007	0.01
	T	0.009	0.005	0.03
	A	0.01	0.01	0.03
	F	0.05	0.04	-
	A	0.02	0.06	-
	S	0.03	0.04	-

Table 39: Comparison of fluency test results to the Hugnarian normative data

It can be concluded that there is a more pronounced difference between the total number of words produced by either groups in our sample compared to the normative sample.

3.3.9. Unique words, genre and word frequency

3.3.9.1. English (FAS)

In the following sections the proportions of the net number of unique words generated by the two groups for the FAS test condition are detailed. The net number of unique words was compared to the total net number of words produced by the two groups for each letter condition. By the 'net number' we mean the single occurrence of a lexical item: a word is counted only once in the total and the repetitions of it are excluded. A 'unique word' is defined as a lexical item produced by (and counted for) one of the groups. 'Overlap' refers to those words that are mentioned by both groups (and for this reason they are not counted for either groups). Table 40 reports on the proportion of the unique and common words:

Net number of words	Control (%)	CLIL (%)	Overlap (common) (%)
F (209)	18.18	40.19	41.62
A (217)	18.89	56.68	24.42
S (351)	23.93	40.45	35.61

Table 40: Proportions of unique and common words in both groups in the English fluency tests

Together the two groups generated 209 different words beginning with letter ‘F’. The control group participants produced 38 unique words (the 18.18% of the total), while the CLIL group participants produced 84 unique words (the 40.19% of the total). 87 words were mentioned by both groups (that is the 41.62% of the total). The two groups generated 217 different words beginning with letter ‘A’. The control group participants produced 41 unique words (the 18.89% of the total), while the CLIL group participants produced 123 unique words (the 56.68% of the total). 53 words were mentioned by both groups (that is the 24.42% of the total). The two groups generated 351 different words beginning with letter ‘S’. The control group participants produced 84 unique words (the 23.93% of the total), while the CLIL group participants produced 142 unique words (the 40.45% of the total). 125 words were mentioned by both groups (that is the 35.61% of the total).

3.3.9.1.1. Genre

In the following sections the proportions of the net number of unique words in different genres generated by each group relative to the total net number of unique solutions are detailed. Table 41 reports on these data:

			Spoken %	Fiction %	Magazine %	Press %	Academic %
UNIQUE WORDS	CLIL	F	0.9	15.31	10.52	4.78	8.61
		A	5.07	13.82	16.59	5.99	15.20
		S	3.42	15.67	10.54	4.55	6.26
	Control	F	0.47	6.22	6.69	2.87	1.91
		A	2.30	4.60	4.60	3.22	4.15
		S	1.99	9.68	6.83	2.27	3.13
Overlap (common)		F	4.30	13.87	11.48	8.13	3.82
		A	3.22	10.59	3.22	0.92	6.45
		S	3.98	17.37	7.97	4.274	1.99

Table 41: Net number of unique words generated by both groups in the different genres in the English fluency tests

Overall, we can conclude that more unique words have been produced in all genres by the CLIL group compared to the control group. Table 42 shows the difference between the groups expressed in percentages reporting on CLIL dominance:

		Spoken (%)	Fiction (%)	Magazine (%)	Newspaper (%)	Academic (%)
Unique words	F	0.43	9.09	3.83	1.91	6.7
	A	2.77	9.22	11.99	2.77	11.05
	S	1.43	5.99	3.71	2.28	3.13

Table 42: Group differences in the various genres expressed in percent in the English fluency tests

A more remarkable difference can be observed in the proportion of words in fiction, magazine and academic language use, while in case of spoken words this difference is slight.

3.3.9.1.2. Word frequency

In the following sections the frequency of unique words generated by the groups are investigated. We defined words as ‘rare’ if they occur less than a thousand times in the COCA corpus. While ‘frequent’ words occur more than fifty-thousand times in COCA. The category ‘inbetween’ refers to those words that are between the two ‘endpoint frequencies’. Table 43 shows the proportions of unique solutions of the two groups (expressed in percentages) relative to the total number of solutions created by them for each letter.

		Frequent (%)			Inbetween (%)			Rare (%)		
		F	A	S	F	A	S	F	A	S
FAS F (209) A (217) S (351)	CLIL	4.78	7.83	6.26	28.23	39.63	2.82	6.22	8.75	6.26
	Control	0.95	3.68	3.7	13.39	12.44	14.81	4.30	2.30	4.84
	Overlap (common)	17.70	11.06	13.96	23.92	12.90	21.65	0.47	1.38	0.28

Table 43: Word frequency in the English fluency tests

Overall, it can be concluded that the CLIL group participants created unique 'frequent' and 'rare' words to a larger proportion for each letter compared the control group. Moreover, relative to each other, the proportions of these categories can be considered balanced. Regarding the proportion of unique 'inbetween' words relative to the total number of unique words is also higher in case of the CLIL group. With regard to the control group, the proportion of 'frequent' and 'rare' categories can be considered as more varied compared to that of the CLIL group.

3.3.9.2. Hungarian (KTA)

In the following sections the proportions of the net number of unique words generated by the two groups for the KTA test condition are detailed. The net number of unique words was compared to the total net number of words produced by the two groups for each letter condition. By the 'net number' we mean the single occurrence of a lexical item: a word is counted only once in the total and the repetitions of it, are excluded. A 'unique word' is defined as a lexical item produced by (and counted for) one of the groups. 'Overlap' refers to those words that are mentioned by both groups (and for this reason they are not counted for either group). Table 44 reports on the proportions of the unique and common words:

All generated words (without repetitions)	Control (%)	CLIL (%)	Overlap (%)
K (409)	33	33.25	33.74
T (392)	32.39	26.02	41.58
A (243)	30.45	34.56	34.98

Table 44: Proportions of unique and common words in both groups in the Hungarian fluency tests

Together the two groups generated 409 different words beginning with letter 'K'. The control group participants produced 135 unique words (the 33% of the total), while the CLIL group participants produced 136 unique words (the 33.25% of the total). 138 words were mentioned by both groups (that is the 33.74% of the total). The two groups generated 392 different words beginning with letter 'T'. The control group participants produced 127 unique words (the 32.39% of the total), while the CLIL group participants produced 102 unique words (the 26.02% of the total). 163 words were mentioned by both groups (that is the 41.58% of the total). The two groups generated 243 different words beginning with letter 'A'. The control group participants produced 74 unique words (the 30.45% of the total), while the CLIL group participants produced 84 unique words (the 34.56% of the total). 85 words were mentioned by both groups (that is the 34.98% of the total).

3.3.9.2.1. Genre

In the following sections the proportions of the net number of unique words in different genres relative to the total net number of unique solutions are detailed.

%			Press	Literature	Academic	Official	Spoken
UNIQUE WORDS	CLIL	K	20.53	7.82	3.17	0.73	2.69
		T	17.09	5.10	1.78	0.76	1.02
		A	22.63	3.29	6.17	1.23	1.23
	Control	K	18.33	9.29	3.42	0	2.20
		T	17.34	5.61	6.63	1.02	1.02
		A	18.93	3.70	5.35	0.82	1.64
Over-lap		K	19.56	8.06	2.69	0	1.47
		T	24.74	8.42	4.33	1.27	3.83
		A	23.87	7.40	2.88	0	0.82

Table 45: Net number of unique words generated by both groups in the different genres in the Hungarian fluency tests

Overall, it can be concluded that the difference between the CLIL and the control group in terms of the number of unique words generated in the KTA (Hungarian fluency) test is less remarkable. However, regarding genres a more robust difference between the groups has been found: CLIL group participants created words related to press and official language use at a higher proportion as compared to the control group participants. Participants of the control group, however, generated words related to literature and academic language use at a higher proportion as compared to the CLIL group participants. Despite the seemingly mixed picture the dominance of the control group in case of the number of unique words in the Hungarian fluency test can be established. Table 46 below shows the difference between the two groups expressed in percentages (numbers in italics indicate the higher performance of the CLIL group, while those in bold indicate that of the control group).

	%	Press	Literature	Aademic	Official	Spoken
Unique words	K	*2.2*	**1.47**	**0.25**	*0.73*	*0.49*
	T	**0.25**	**0.51**	**4.85**	**0.26**	0
	A	*3.7*	**0.41**	*0.82*	*0.41*	**0.41**

Table 46: Group differences in the various genres expressed in percent in the Hungarian fluency tests

3.3.9.2.2. Word frequency

Finally, we investigated the occurrence rate of the rarest and most frequent words generated by the groups. Words occurring less than a thousand times have been defined as the rarest words, while those occurring fifty-thousand or more times have been defined as the most frequent ones based on the MNSZ2 corpus. Table 47 shows the proportions of unique solutions relative to the total number of solutions produced by the groups per letter.

Unique words		Frequent (%)			Inbetween (%)			Rare (%)		
		K	T	A	K	T	A	K	T	A
KTA K=409 T=392 A=243	CLIL	0.24	**0.51**	1.64	**7.33**	6.63	**6.99**	28.36	18.87	**25.92**
	Control	0.24	0	**2.47**	7.09	**15.81**	6.17	28.36	**26.78**	21.81
	Common	0.73	0.51	4.12	5.38	5.61	15.22	22.25	25.25	15.64

Table 47: Word frequency in the Hungarian fluency tests

Overall, it can be stated that the indicators show a similarly mixed picture in terms of the proportion of the frequent-rare words in case of both groups. The same can be observed for the proportion of unique ’inbetween’ words. Finally, Table 48 and 49 summarize those variables in which one of the groups had more dominant results in comparison to the other. Highlighting indicates twice as higher results compared to the other group’s results.

		F	A	S
Unique words	Spoken	CLIL	CLIL	CLIL
	Fiction	**CLIL**	**CLIL**	**CLIL**
	Magazine	**CLIL**	**CLIL**	**CLIL**
	Press	CLIL	CLIL	CLIL
	Academic	**CLIL**	**CLIL**	**CLIL**
	Frequent	**CLIL**	**CLIL**	**CLIL**
	Rare	CLIL	**CLIL**	CLIL

Table 48: Summary table on group dominance in the English tests for genre and frequency

		K	T	A
Unique words	Press	CLIL	Control	CLIL
	Literature	Control	Control	Control
	Academic	Control	**Control**	CLIL
	Official	CLIL	Control	CLIL
	Spoken	CLIL	-	Control
	Frequent	-	CLIL	Control
	Rare	-	Control	CLIL

Table 49: Summary table on group dominance in the Hungarian tests for genre and frequency

3.4. Qualitative Content Analysis

While analyzing the interviews we focused on the following indicators: signs of mental flexibility (the fluency of ideas, complexity of ideas, expression of abstract thoughts, specific approaches to the issues, shifting among thoughts) and the quality of utterances (style and length).

Based on the results of the quantitative study twelve learners (N=12) have been selected for the qualitative study. These learners were classified into four groups: CLIL high (N=3), CLIL low (N=3), control high (N=3) and control low (N=3). Selected learrners' results and characteristics can be seen in Appendices I-L. This his way four document-sets formed the basis of the qualitative content analysis. The questions of the structured interview were arranged along the following seven main codes: aim of learning, language teacher, lessons, activities, competences, development and do differently. These main codes represent the larger content units of the interview, that can be broken into sub-codes in accordance with the textual content. Sub-codes are considered as signs of fluency in ideas. These code types form the basis of the analysis along which high score groups have been compared to low score groups. Table 50 shows the aspects of the analysis:

Main codes	Answering groups
Aim of Learning	CLIL High – Control High; CLIL Low – Control Low
Language teacher	
Lessons	
Activities	
Competences	
Development	
Do differently	

Table 50: Main codes and the groups involved in the interviews

According to MAXQDA project information, 132 codes were generated in 413 coded units of the 12 interviews.

3.4.1. Aim of learning (‘high’ groups)

In terms of questions related to English language (What is the most important in learning English? Why are you learning English?) CLIL high learners provided more and more varied answers than the control group learners. While the control group learners described the importance of L2 learning from a more ‘skill-based’ aspect (laying the emphasis on communication, vocabulary, comprehension and grammar), CLIL group learners added two more (pronunciation and confident language use) to this list; from which the latter one suggests a more holistic and abstract view on L2 learning (‘*Az a legfontosabb az angoltanulásban, hogy bátran és magabiztosan beszéljünk, mert néhány nyelvi hibával is megértenek, és jobb mintha egyáltalán nem szólalunk meg.*’ Regarding the reason of L2 learning the answers are even more varied for the CLIL group learners. They not only listed more (7) answers but five of them were unique as well: they find it crucial to be understood if they are abroad, they also set the goal of taking the intermediate language exam (‘*...Látom, milyen nehéz az, ha külföldön vagy és nem tudod megértetni magad*’). Some of them found the L2 interesting and motivating. This way a much more positive attitude emerges from CLIL learners’ responses.

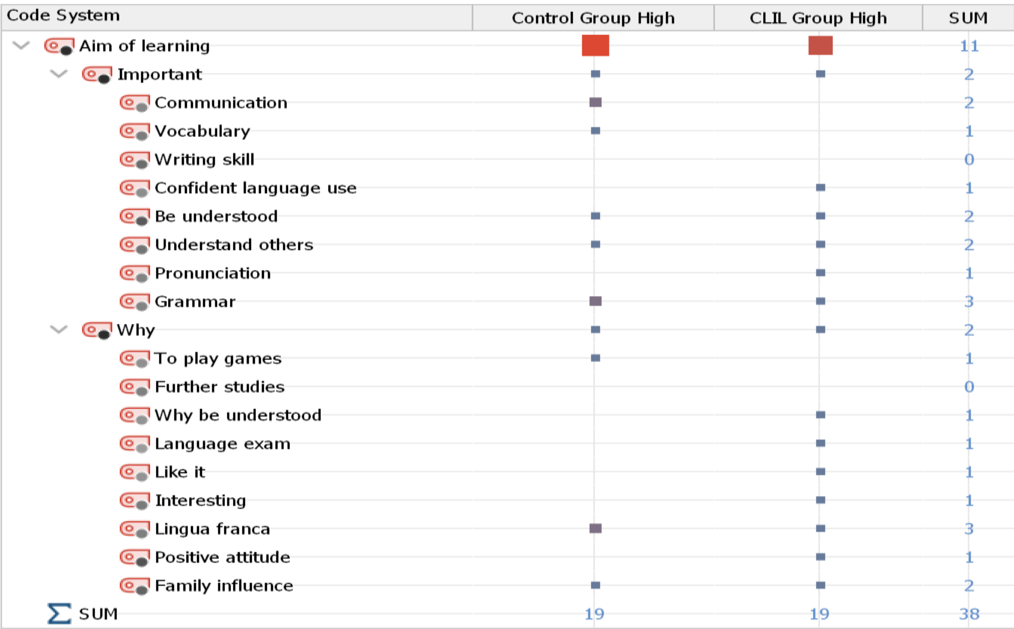


Figure 13: Aim of L2 learning (‘high’ groups)

In both groups, English is mentioned as a lingua franca as the reason and aim of L2 learning (‘*...Az angolt a letöbb helyen beszélik...*’, ‘*...ez ma már alapvető...*’) Given that, learners are surrounded by the L2 in their daily lives, this reflection does not seem to be an unusual answer. Some learners in both groups mentioned family influence as the reason of L2 choice (*...a keresztanyám ötlete volt, hogy angolul tanuljak...*’, ‘*... a bátyám ebben a nyelvben tud segíteni...*’). However, one of the control group members provided an unconventional reason of L2 choice: ‘*A legtöbb videojáték csak angol nyelven elérhető*’. Figure 13shows ‘high’ learners’ answers of the related questions on the aim of L2 learning.

3.4.2. Aim of learning (‘low’ groups)

Generally, it can be stated that the ‘low’ groups’ reflections are not as complex as those of the ‘high’ groups. This conclusion is confirmed by the MAXQDA statistics (sub-codes that have a zero score were not mentioned among their answers at all). Figure 14 shows ‘low’ learners’ answers of the related questions on the aim of L2 learning.

Figure 14: Aim of L2 learning (‘low’ groups)

Both groups mentioned ‘further studies’ and ‘lingua franca’ as the main reasons of their L2 learning: ‘*Az informatika érdekel, és ahhoz szükségem van az angolra*’, ‘*Fontos, hogy tudd a nyelvet, mert az egész világon ezt használják*’.

‘Family influence’ and ‘interest’ only appear in case of the control group: ‘*Első óta ezt tanulom, ezt választották nekem*’. However, this last comment does not suggest a positive attitude towards L2 learning. ‘Pronunciation’ is mentioned by both groups, however a CLIL group participant approaches it from an unusual point of view: ’*Szerintem a kiejtés a legfontosabb az angoltanulásban. Ha ismersz egy szót, de nem tudod kiejteni, annak semmi értelme.*’ ‘Vocabulary’ and ‘writing skill’ appeared as extras in case of CLIL learners, although control group learners mentioned the ‘understanding of others’ as an additional idea.

3.4.3. Characteristics of a good language teacher (‘high’ groups)

The topic of the language teacher is investigated along the sub-codes ‘good language teacher’ and ‘relationship with the language teacher’. (In case of the CLIL group all subject teachers teaching in the L2 were regarded as language teachers.). Figure 15 shows how ‘high’ learners describe a ‘good language teacher’.

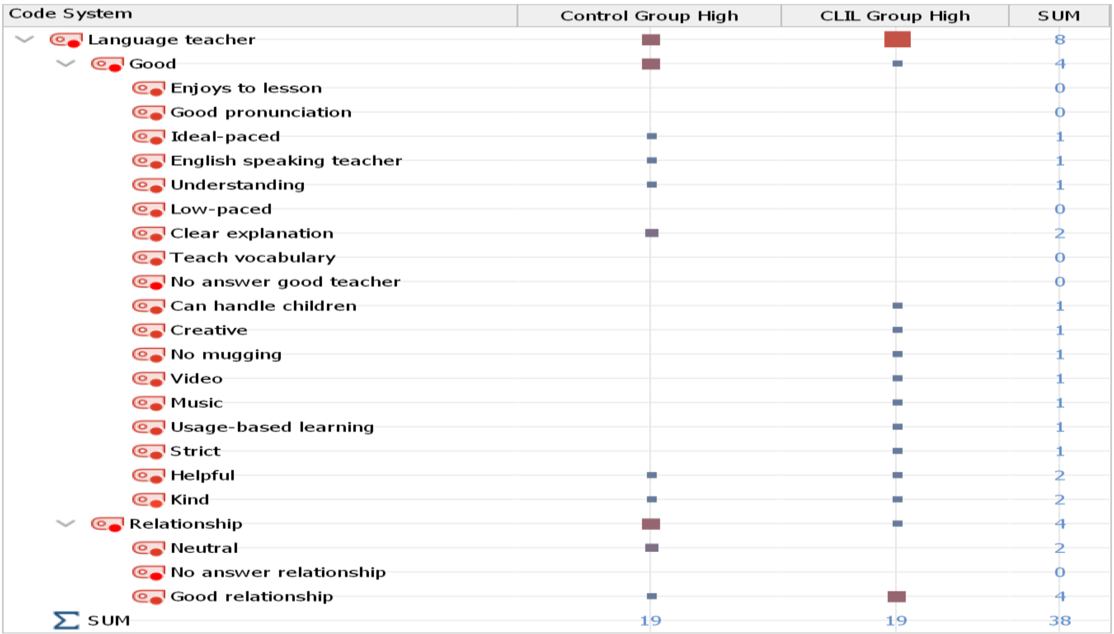


Figure 15: Characteristics of a good language teacher (‘high’ groups)

In case of the CLIL ‘high’ group the complexity and diversity of the reflections can be clearly seen compared to the control group. Their visions and expectations of a good language teacher’s characteristics and practice presuppose the existence of different teaching methods in the lessons.

Moreover, the ability to perceive these from a more holistic perspective, assumes learners' ability to think over the questions critically. Control group 'high' participants approached the issue along two clear dimensions that is the teacher's personality and class managing techniques ('*Ha valamit nem értünk, elmagyarázzák.*', '*Angolul beszélnek az órán*', *'...aki nem dedós tempóban tanít, de nem is sieti el...'*.

In contrast CLIL 'high' group participants' views touched upon at least three dimensions regarding the issue: a good teacher's main characteristics ('*szigorú és segítőkész*), behaviour ('*...tud bánni a gyerekekkel...*') and teaching methods. This latter one however can be subdivided into three more dimensions: a holistic view on L2 learning ('*...az órán használjuk a nyelvet..*'), a view on teacher's methodological culture ('*...zenét hallgatunk és videót nézünk...*') and students' learning strategy ('*...nem kell magolni a nyelvtant...*'). This flow of ideas does not only confirm the ability of breaking away from reality, fast retrieval, critical thinking but shifting among these ideas as well.

The question related to the teacher-learner relationship aimed to reveal those background motives that might have influence on learners' general opinions on the issues. Since the interviews were also completed anonymously, learners could feel free to express their real opinions. The teacher-learner relationship was much more positively worded by the CLIL participants ('*...jól kijövünk...*') compared to the control group participants who assessed it as neutral ('*Semleges.*', '*Nem rossz'*).

3.4.4. Characteristics of a good language teacher ('low' groups)

In case of the 'low' groups, answers are mainly related to basic teaching methods. Control group learners' answers, however, can be divided into four main categories (as opposed to the two provided by the CLIL group): attitude (e.g. teacher is required to enjoy the lesson), teacher's L2 ability (e.g. has a good pronunciation), teaching method (e.g. explain the material clearly) and characteristic (e.g. helpful). Since these ideas only superficially illustrate the language teacher and the teaching methods, we can conclude two things: a) control group participants are not taught in a methodologically-rich language environment or b) if they are, they are not conscious enough to be able to perceive its constituents. It is also worth noting that the same two learners (one in each 'low' group) did not provide answers for the questions related to the good teacher characteristics and the teacher-learner relationship. Figure 16 reports on 'low' groups' answers:

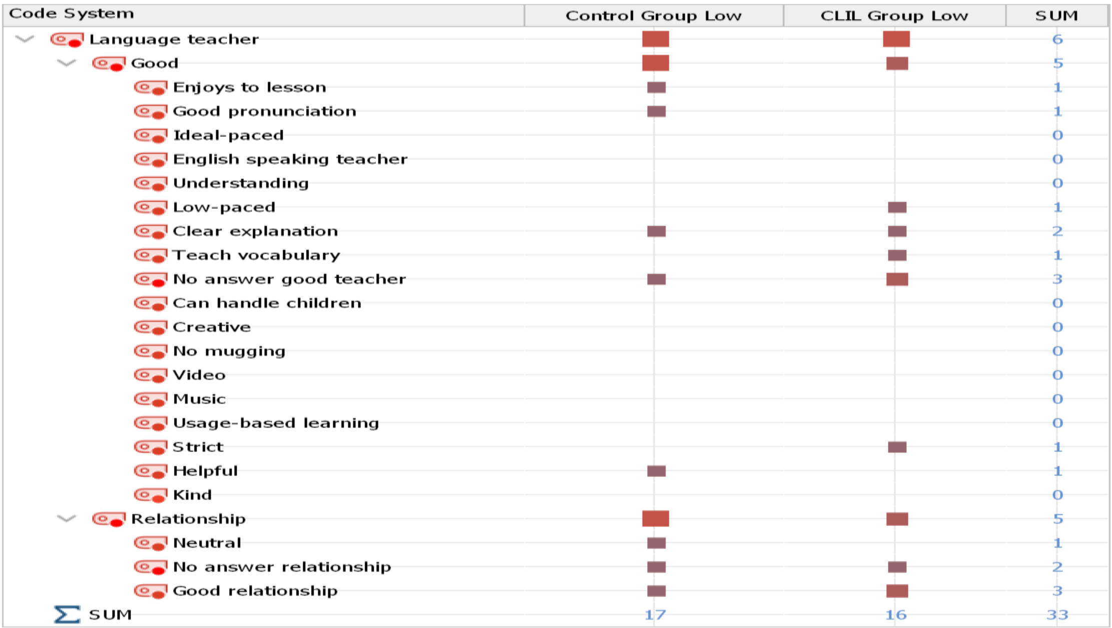


Figure 16: Characteristics of a good language teacher ('low' groups)

3.4.5. L2 lessons or lessons in the L2 ('high' groups)

In the next sections learners' attitudes towards the quality of their lessons are investigated. In general we can conclude that CLIL learners have a more positive attitude regarding the L2 lessons or lessons in the L2. Their reflections are versatile as assumed: they report on fast-paced, funny, creative lessons in which they do different activities like chatting, listening to stories watching films and doing playful tasks ('*A legtöbb óra élvezetes, mert sokszínű.*', '*Azokat az órákat kedvelem, amikor a tanárnak sikerül valamilyen story-t becsempésznie az óra menetébe*').

Two learners in the control group, however, approached the same issue from an unexpected perspective: they emphasize the importance of the company whom they have to cooperate with. They claim that it is primarily the company and the atmosphere that make a lesson good not the contents of it. This opinion suggests a joy-managed rather than goal-directed behaviour. ('*Élvezem, mert jó a társaság.*', '*Attól függ, hogy mi a téma, milyen a légkör és kikkel kell együtt tanulnom.*' Figure 17 shows the answers related to the quality of the lesson:

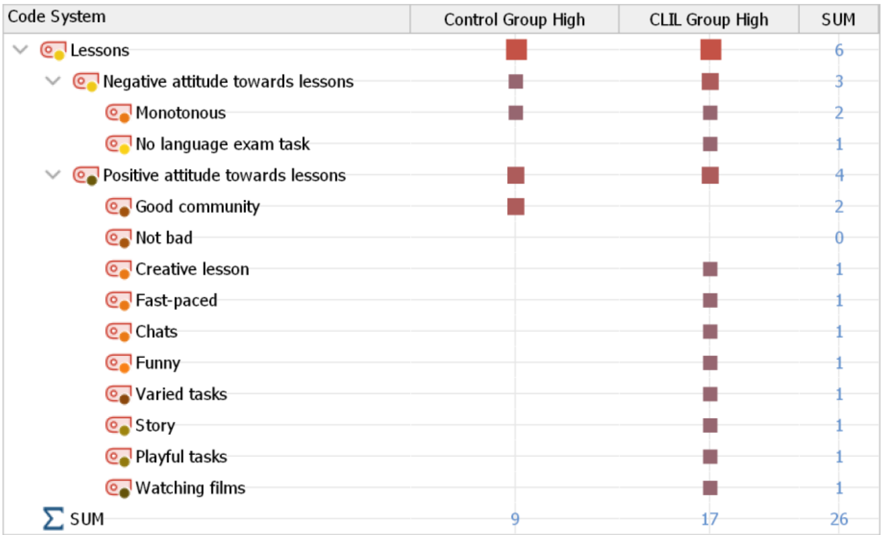


Figure 17: L2 lessons or lessons in the L2 (‘high’ groups)

Negative comments on the quality of the lessons are made in three cases: a control group learner does not enjoy the lessons at all, since no playful activities are involved; a CLIL group learner finds History demanding and boring because of the lack of proper understanding. Another CLIL group participant assesses language exam tasks ‘a factor’ that has a detrimental effect on the quality of L2 lessons.

3.4.6. L2 lessons or lessons in the L2 (‘low’ groups)

In case of the ‘low’ groups, there is a high consistency among the opinions and the less positive attitude towards the L2 lessons or lessons in the L2. However, CLIL group participants report on lessons somewhat more positively: they appreciate the varied tasks, although sometimes there are too many of them (‘We always have tasks to do in the English class, sometimes too many. Otherwise, the other lessons are boring’). One of them claims that the lessons are not bad, if they interest them. Control group participants, however, do not provide an appreciable opinion while saying ‘*Szerintem jók, de lehetnének jobbak is’*. Regarding negative attitudes control group participants expressed their definitely pronounced opinions: ‘*Jobb lenne, ha nem cask állandóan írnánk és interaktívabb lenne, kérdeznének minket, mi meg válaszolnánk.*’, ‘*Gyakran unalmasak és lassúak’*. Figure 18 reports on the answers of the ‘low groups’:

Figure 18: L2 lessons or lessons in the L2 (‘low’ groups)

3.4.7. Activities (‘high’ groups)

The opinions related to activities in the lessons are investigated via popular and unpopular tasks. In terms of popular tasks and activities, the dominance of the CLIL group can be detected, since its participants produced more unique sub-codes compared to the control group. These sub-codes are as follows: creative tasks, listening comprehension, listening to music, watching videos, essay writing, text completion and reading. Control group members only list two activities (crosswords and playful tasks). Regarding the disliked activities, again, CLIL (‘high’) group participants listed more task types: word search, listening, speaking, writing and crosswords. Control group participants also listed listening among the least preferred tasks, just like reading-translation activities. Figure 19 shows the ‘high’ groups’ reflections on preferred and disliked task types:

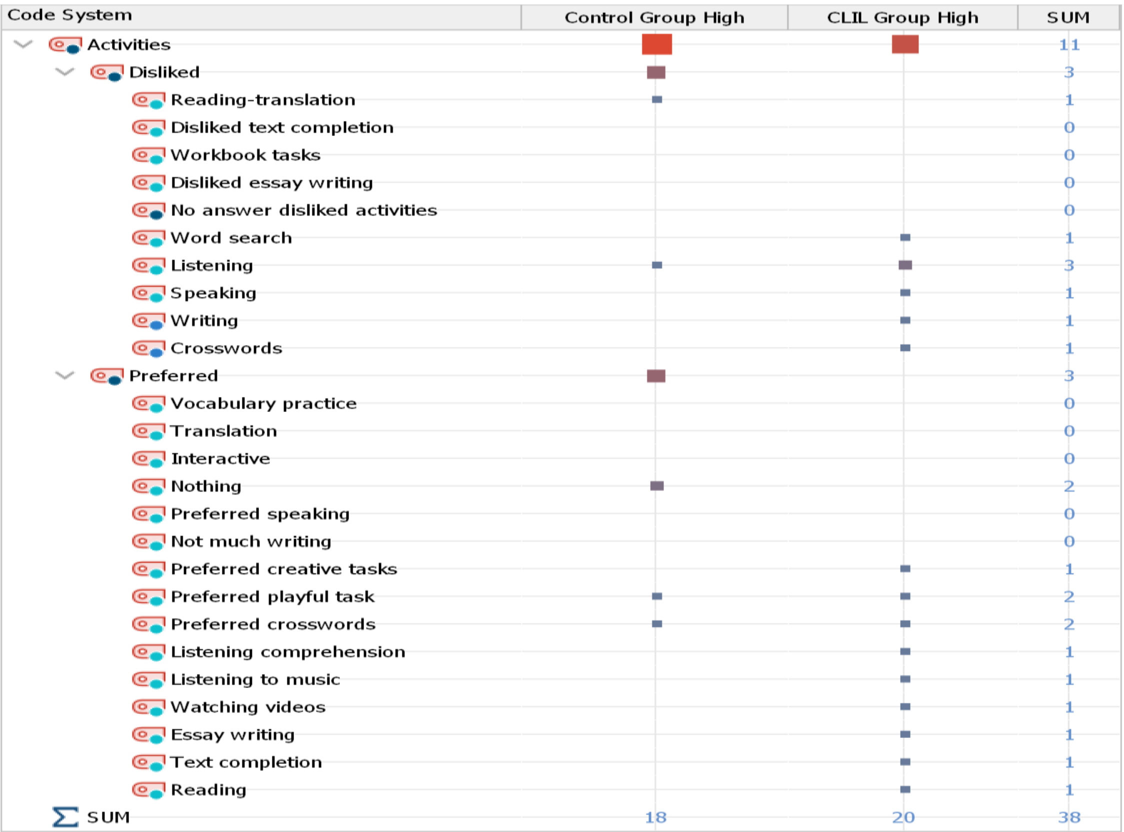


Figure 19: 'High' groups' reflections on preferred and disliked activities

3.4.8. Activities ('low' groups)

In general it can be concluded that the proportion of reflections on preferred and disliked activities produced by the 'low' groups is similar to that of the high groups: CLIL group participants provided more and more diverse reflections for both questions. They listed 'old-school' methods (e.g. text completion, workbook activities and essay writing) among the disliked activity types and so did control group participants (reading-and-translating and writing). In case of the CLIL groups ('low' and 'high') there is a remarkable similarity in terms of preferred activities: learners in both groups list creative, playful tasks, crosswords and videos among them. Regarding preferred activities control group participants' reflections are less diverse and seem to got stuck at the level of classical methods again (reading, vocabulary practice, translation). (Reading is defined as their strength in the L2 in the LEAP-Q questionnaire.) Figure 20 reports on the reflections on preferred and disliked activities given by both 'low' groups:

Figure 20: ‘Low’ groups reflections on preferred and disliked activities

3.4.9. Development (‘high’ groups)

CLIL learners’ tendentiously more complex and diverse approaches can also be detected on the issue of their L2 development Figure 21 reports on ‘high’ learners’ L2 development:

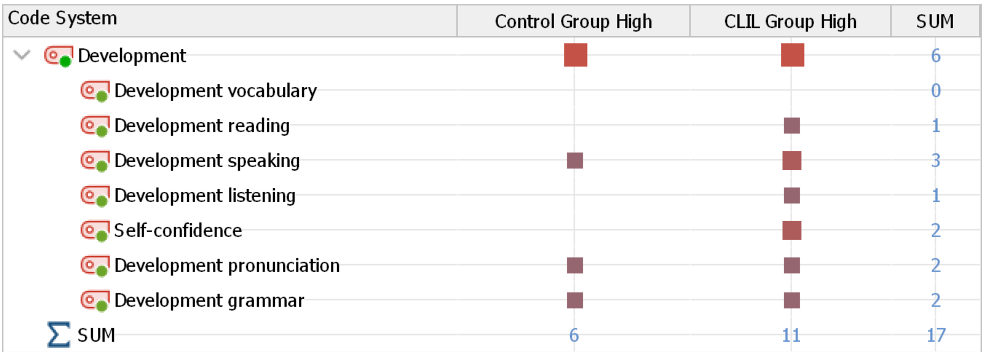


Figure 21: ‘High’ learners’ L2 development

Two of the CLIL learners, however, highlight their development in speaking. In addition, two of them confirm their developing confidence also (‘*Még mindig nem vagyok elég magabiztos, ha beszélek valakivel.*’, ‘*Sokkal magabiztosabb vagyok, mint korábban, de még mindig vissza kell kérdeznem, ha valamit nem értek’*). In contrast, control group participants listed only three L2 components (e.g. speaking, pronunciation and grammar) that can be categorized into the same group. Reflecting to their own learning paths needs the ability to change perspectives that seems problematic in their case.

3.4.10. Development (‘low’ groups)

In terms of L2 development control group participants provided more complex and diverse reflections covering some L2 skills (e.g. reading, speaking, and pronunciation) and personal characteristics (self-confidence). In contrast, CLIL group participants, only listed listening and vocabulary. Surprisingly, no overlap can be detected between the reflections of the two groups; that is clearly displayed in Figure 22:

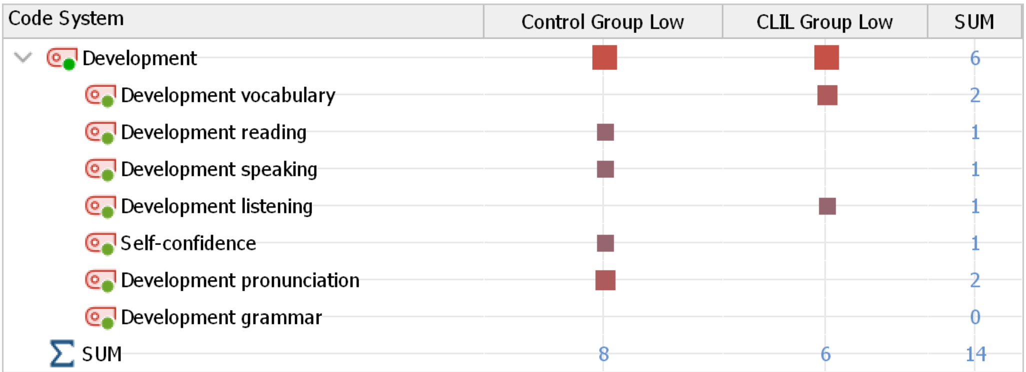


Figure 22: ‘Low’ learners’ L2 development

3.4.11. Competences (‘high’ groups)

Under the key word ‘competences’ two related questions have been recruited (What are your strengths in learning English? What are your weaknesses in learning English?). Regarding the reflections to these questions, the previous trend seems to continue: CLIL group participants provided more and more diverse reflections especially in case of strengths.

In addition to the classical didactic elements of an L2 lesson (e.g. writing, listening comprehension and pronunciation), the desire for expressing opinion (as a unique answer) also appears in this group (‘*Szeretem elmondani a véleményem*’). Speaking (‘*Könnyen el tudom mondani, amit szeretnék*) and memory (‘*Könnyen megjegyzek dolgokat*’) appear among the reflections of both ‘high’ groups, while reading and learning vocabulary (as more conventional answers) are only mentioned by the control group participants. In terms of weaknesses CLIL group participants have also been more critical: all of them considered grammar as a crucial point in their L2 knowledge. They also identified ‘asking words’ and ‘listening comprehension’ as weak points. Given that, CLIL learners are constantly encouraged to paraphrase if they do not know the exact meaning of a word; it is not surprising that ‘asking words’ is inconvenient for them. So is listening comprehension, which is also a task in the L2 exam and is linked to the oral part of it. There might be pressure on learners, because if they fail in the listening comprehension (or do not achieve the required minimum scores) this exam part should be performed again. If CLIL learners’ reflections on their assumed strengths and weaknesses are considered entirely, it can be concluded that their remarks are strongly related to the requirements of an intermediate L2 exam. Figure 23 reports on ‘high’ group participants’ strengths and weaknesses:

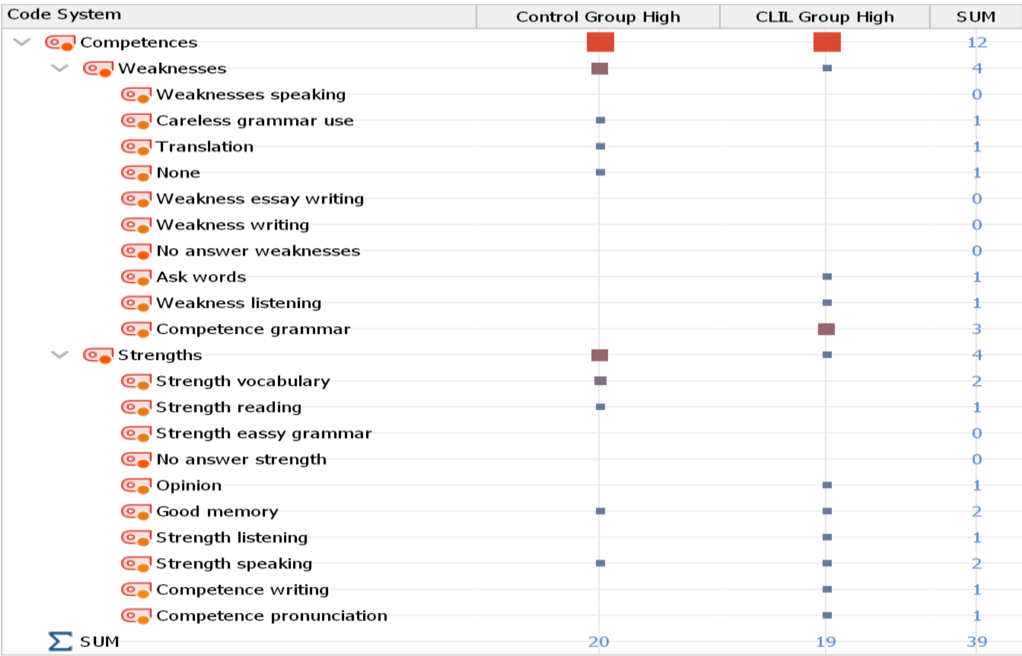
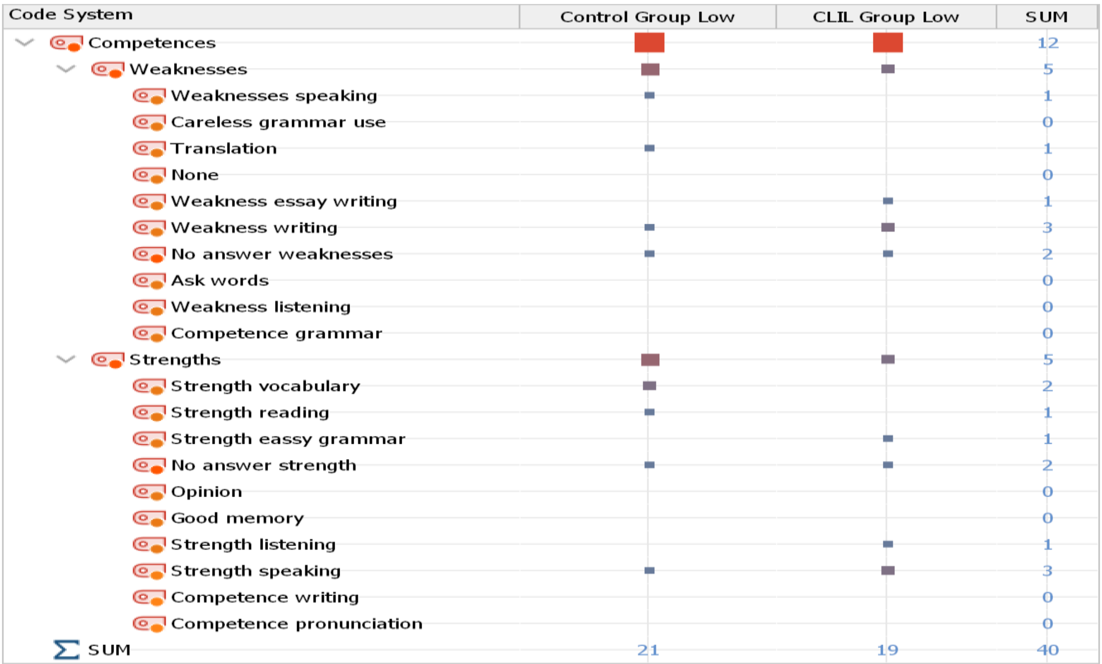


Figure 23: Strengths and weaknesses (‘high’ groups)

In terms of weaknesses, control group participants seem to be unable to approach the question from a different perspective: grammar and translation are dominant in their reflections. One of them does not have an idea at all (‘*Nincsenek’*.) suggesting the lack of critical thinking or that of motivation to give a detailed answer.

3.4.12. Competences (‘low’ groups)

In ‘low’ groups, there is also a tendency for a more balanced number and type of reflections. However, overlap between the two groups cannot be observed, they are common in at least three things: a) leaving some questions unanswered, b) providing less answers than their ‘high’ counterparts and c) repeating their previous answers (e.g. translation and writing). However, ‘low’ CLIL learners’ have the additional view point of essay writing (related to the language exam again). Conversely, ‘low’ control groups cannot list any novel ideas among their weaknesses. Figure 24 lists on ‘low’ group participants strengths and weaknesses:



Code System	Control Group Low	CLIL Group Low	SUM
Competences			12
Weaknesses			5
Weaknesses speaking			1
Careless grammar use			0
Translation			1
None			0
Weakness essay writing			1
Weakness writing			3
No answer weaknesses			2
Ask words			0
Weakness listening			0
Competence grammar			0
Strengths			5
Strength vocabulary			2
Strength reading			1
Strength eassy grammar			1
No answer strength			2
Opinion			0
Good memory			0
Strength listening			1
Strength speaking			3
Competence writing			0
Competence pronunciation			0
SUM	21	19	40

Figure 24: Strengths and weaknesses (‘low’ groups)

3.4.13. Do differently (‘high’ groups)

The last question (If you were an English teacher, what would you do differently in the lesson?) requires creativity, shifting among ideas and the ability of breaking away from the conventional solutions.

Again, CLIL group participants provide more responses, some of which have not been listed before like chatting, little homework, more cooperation with peers and dual-language explanation. Some of their answers are related to teachers' classroom management (dual-language explanation, little homework, learning through games), while others suggest their desire for being more actively involved in the activities (more groupwork, chatting). Control group learners' responses, however, are less specific in this question, and most of them have been listed before (videos and films, learning through games, less teaching material) and are learner-centred. Their remarks suggest that they would put less effort in L2 learning. Figure 25 reports on 'high' groups' remarks on what they would do differently in the L2 classroom:

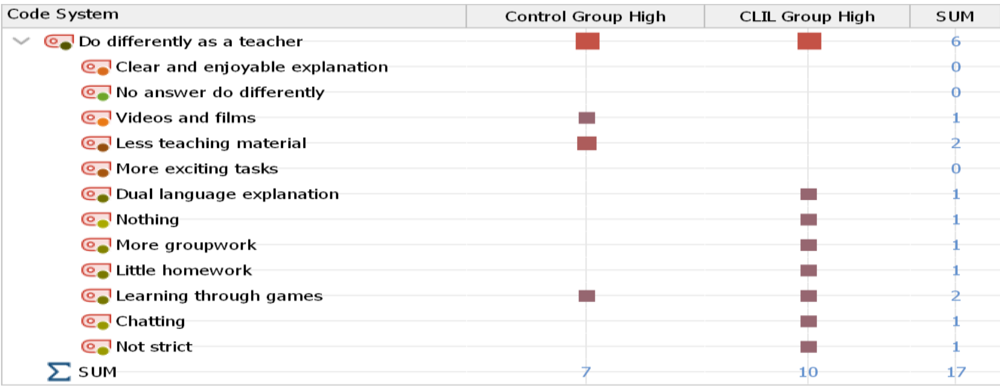


Figure 25: 'High' groups' reflections on what learners would do differently in the L2 lesson

3.4.14. Do differently ('low' groups)

Responses related to the differences that 'low' group learners would make in the L2 lesson, a more balanced situation is outlined. However, they listed less sub-codes than the 'high' groups. Figure 26 reports on 'low' groups' remarks on what they would do differently in the L2 classroom:

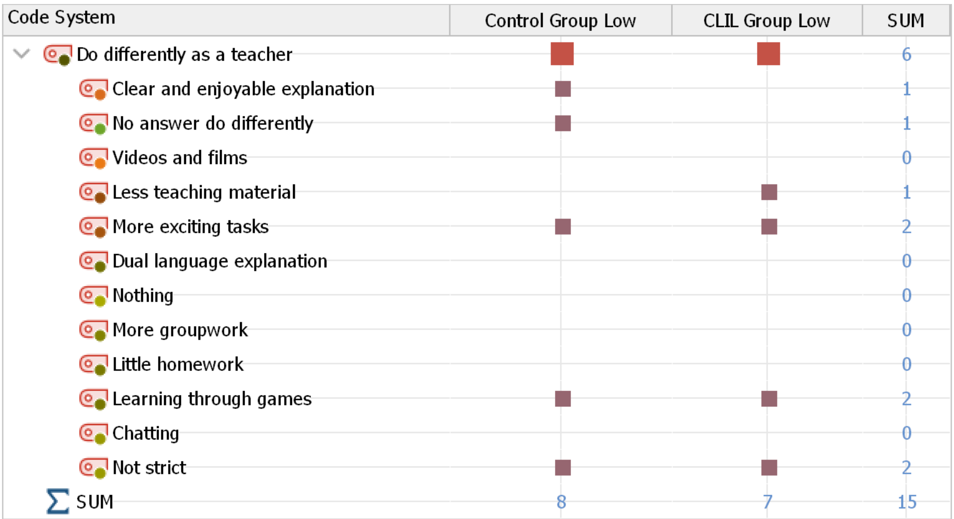


Figure 26: 'Low' groups' reflections on what learners would do differently in the L2 lesson

The investigation of the lists provided by the two 'low' groups reveals that learners would like to settle in a comfort-safety position and be entertained by the teacher in the lesson without taking an active part in it ('*Értelmesen és beleéléssel magyaráznám az anyagot.*').

The following figures (created via MAXQDA) display the code hierarchy of the interviews provided by the 'high' groups. Observing the structure and the number of components of the figures, it can be concluded that the CLIL 'high' group participants had the highest fluency of ideas among the four groups They created 68 different sub-codes, nearly twice as many as the control 'high' group participants (36). The highest number (9) of sub-codes created for a main category by this group is related to the characteristics of a good language teacher and the preferred activities. The lowest number (1) of sub-codes is related to their relationship with their teachers. Figure 27 reports on the code hierarchy of the CLIL 'high' group:

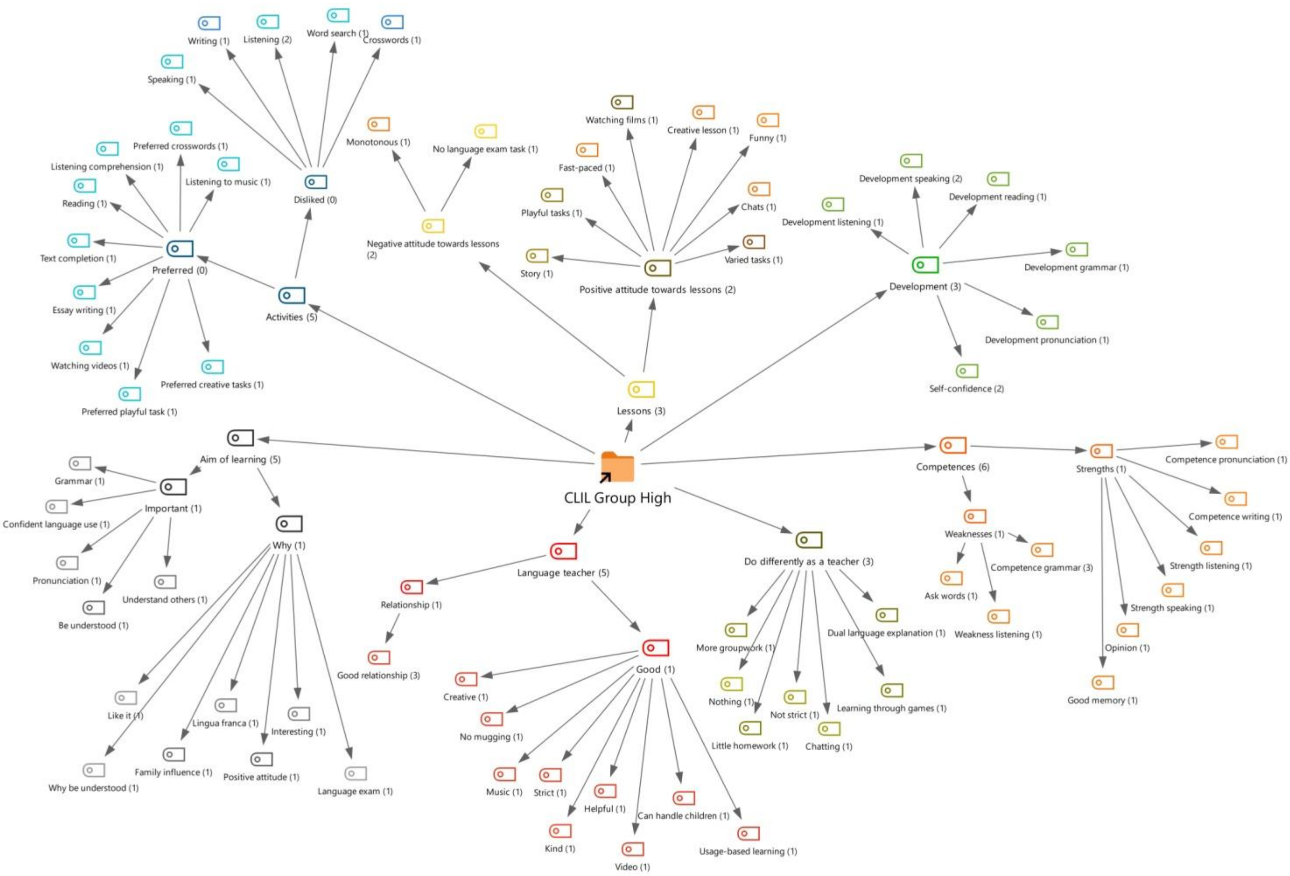


Figure 27: Code hierarchy in case of the CLIL ‘high’ group

In general we can conclude that CLIL ‘high’ group participants retrieved ideas with ease and were able to approach the same questions from many and often various perspectives which is reflected in the number of created categories. The number of unique ideas (35) in the categories is also remarkable as compared to that of the control ‘high’ group (11).

Conversely, control ‘high’ group participants have not gone to the extremes in terms of the number of sub-codes. With their 36 sub-codes they take the middle position along with the control ‘low’ group. In case of the control ‘high’ group the most remarkable number of sub-codes is 6 (related to the characteristics of a good language teacher) and the least is 2 (related to the attitudes towards the L2 lessons). Figure 28 reports on the code hierarchy of the control ‘high’ group:

Figure 28: Code hierarchy in case of the control 'high' group

Generally, it can be concluded that the number of responses provided by the control 'high' participants is balanced.

Observing the structure and the number of components of the figures of the 'low' groups, we can conclude that the CLIL 'low' group participants and control 'low' participants have created nearly the same number of sub-the codes (ideas). CLIL participants created 35 different sub-codes of which 11 were unique in their categories. Similarly, control 'low' group participants created 37 sub-codes of which 7 were unique. The highest number (9) of sub-codes created for a main category by the CLIL 'low' group is 5 and related to the characteristics of a good language teacher and the preferred activities. The lowest number (1) of sub-codes is related to their relationship with their teachers. Figure 29 reports on the code hierarchy of the CLIL 'low' group:

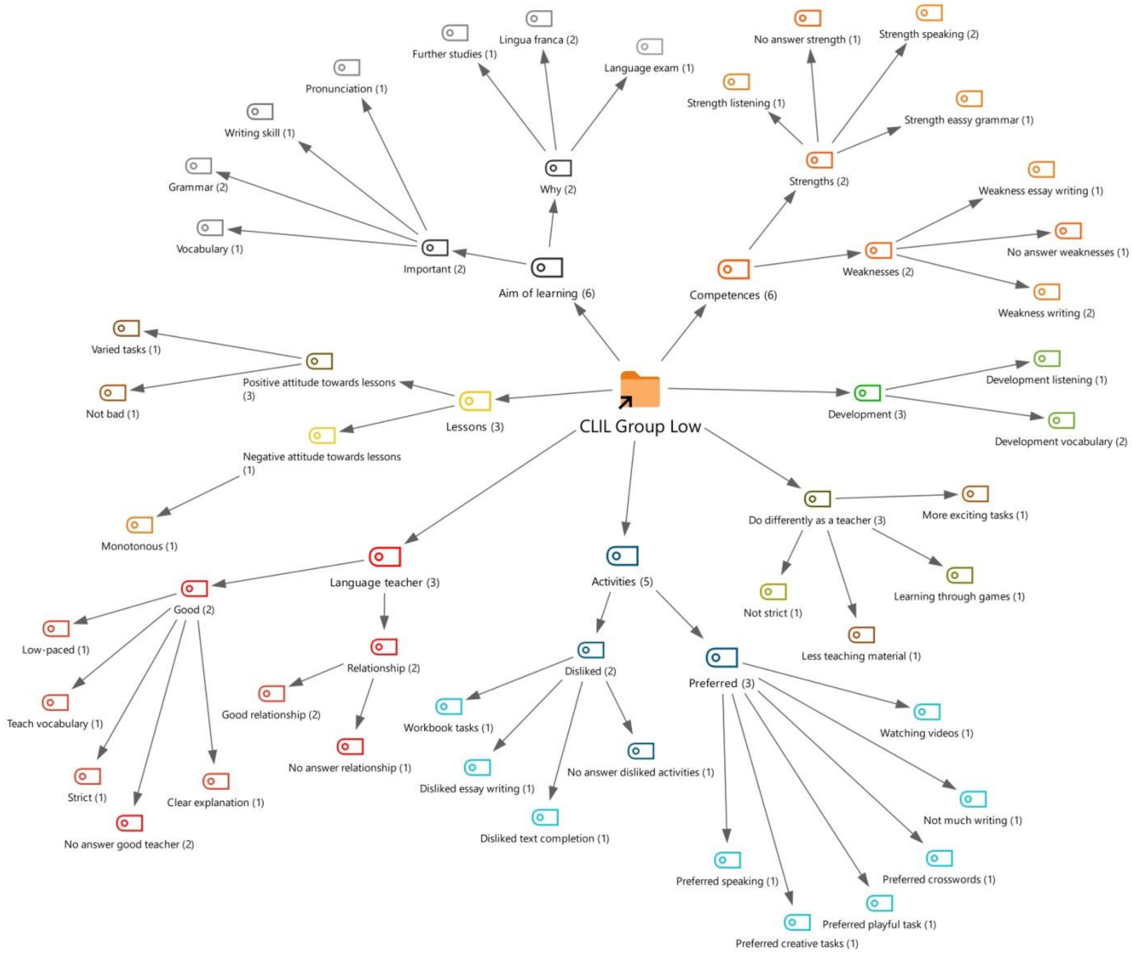


Figure 29: Code hierarchy in case of the CLIL 'low' group

The highest number of sub-codes created for a main category by the control 'low' group is 4 and related to the preferred activities. The lowest number (2) of sub-codes is related to their relationship with their teachers and their attitudes towards the L2 lessons. Figure 30 reports on the code hierarchy of the control 'low' group:

Figure 30: Code hierarchy in case of the control ‘low’ group

3.4.15. Differences in wording between the CLIL and the control groups

Differences in the formulation of learners’ opinions are also worth investigating. While CLIL learners’ statements were formulated in a stylistically expected manner, some control group learners used word phrases that did not match the style of the topic articulated in the question (*‘...aki nem dedós tempóban tanít, de nem is sieti el...’; Sokszor figyelmetlenül, csak ‘csuklóból’ írom a választ.’*). Conversely, CLIL learners, even in the ‘low’ groups, often try to use unique phrases to express their opinions (*‘Azokat az órákat kedvelem, amikor a tanárnak sikerül valamilyen story*-*t becsempésznie az óra menetébe’*, *‘Inkább megkérdezek egy szót, ahelyett, hogy körülírnám, vagy szinonimát keresnék., Olyan nyelvtani feladat [az erőssége]*, *amely könnyű, de pontos szabályismeretet kér.’*)

CLIL learners generally express their opinions in longer and structurally more complex sentences (*'Az angol nyelvű történelem számomra nagyon unalmas, mert nem mindig értem.')* compared to the control group learners (*'Semennyire sem. Jó lenne, ha játékos lenne.', 'Mondatok megfogalmazása.'*). CLIL learners generally use more adjectives than their control counterparts ('*Többnyire élvezetesek mert változatosak, de van, amikor a tanárok nem igazán találják el, hogy mi érdekel minket. A vicces, színes, beszélgetős órák élvezetesek.*', '*Szeretnék több nyelven beszélni, mert az angol egy elterjedt nyelv.*', '*Elnézőbb lennék a diákokkal.*')

3.4.16. Unanswered questions

If the number of missed (unanswered) questions is regarded, the lack of goal-oriented behaviour can be detected again. While both 'high' group participants answered all the questions, some 'low' group participants missed a few of them: six questions remained unanswered by the control group and seven by the CLIL group. To answer these questions (questions 3, 4, 7, 8, 9, 11) learners need to shift perspectives (reflecting on their own and then the teachers' characteristics; reflecting on the present then shift to a longer time period).

CHAPTER 4. Discussion

In this chapter I interpret and contrast the results of the applied questionnaire and the different test types in both groups. To my knowledge this is the first study that aims at comparing executive functioning and verbal abilities of Hungarian CLIL learners and traditional learners of English using Mixed Methods.

4.1. LEAP-Q

The application of LEAP-Q aimed at revealing learners' exposure to L1 and L2 in and out of school. In the research, 142 participants reported Hungarian as their L1. Only three of them named a language other than English as their L2. In both groups, the amount of time spent in the language environment in case of the L2 is remarkably lower compared to that of the L1.

Regarding the presence of the languages in their everyday lives, it can be concluded that CLIL learners have a more positive attitude towards using English for different purposes than the control group participants. This positive attitude might be supported by the motivating classroom activities that feature CLIL methodology and lead to more motivated learners (Lasagabaster, 2008; Coyle et al., 2012). The difference between the assumed frequencies of language use in case of both languages regarding the groups is remarkable. In the case of control group learners, the difference between language use in L1 and L2 is more than twice as much as in the experimental group, whose members seem to be more open to use both languages. It can be stated that CLIL learners, in general, would chose both languages for reading and listening nearly at the same rate. However, they would choose L1 over L2 at a higher rate for talking with somebody. While they would choose both languages for reading with almost equal frequency, this cannot be stated about the control group participants, who definitely prefer reading in L1 to L2. In terms of talking with someone, control group learners seem to consciously choose L1 as opposed to L2. The findings related to the CLIL group are in line with the main objectives declared for the programmes in the MKM decree and the Ministerial Decree 5/2020. (I. 31.).

Learners' language preferences, in general, are in accordance with the self-assessed level of their language skills. The main difference in terms of language skills between the two groups is their assumed level of L2, while they report on nearly similar L1 proficiency. These findings confirm previous research outcomes on the extensive L2 learning. In the CLIL group, the higher levels of self-assessed L2 receptive skills and the estimated lower-level productive skills are in line with previous empirical findings (Pérez-Cañado, 2012; Dallinger et al., 2016). In L1, CLIL learners assume weaker skills in writing and speaking and find listening and reading as the least problematic skills. Since in CLIL programmes the development of productive skills is a primary objective, these results might seem contradictory. However, they might only reflect the meaning-making process that characterizes naturalistic language acquisition, and not improper CLIL methodology (Larsen-Freeman, 2012; Kovács, 2009; Goris et al., 2019; Escobar-Urmenta, 2019). Pérez-Cañado (2012) posited writing, syntax, nontechnical language, and pronunciation are the least affected by CLIL methodology, since there is insufficient focus on form. In L2, speaking is an integrative part of CLIL programmes; hence it comes more naturally and invisibly in comparison to writing (Mehisto et al., 2009; Attard Montalto et al., 2016; Borowiak, 2019). Writing needs a more explicit way of teaching and a more analytic way of learning from students. Considering that primary school learners are more attuned to implicit teaching-learning methods (Pfenninger & Singleton, 2019), learning to write properly might seem more rule-governed than speaking. For this reason, they might not learn it with that ease, and they might need extra effort to develop it giving them the impression of being less skilled in it. This assumption is also confirmed by the fact that they found it the most dominant skill in language learning at the time of data collection.

Given that both CLIL and control group learners considered reading as their least problematic skills in the L2, we might assume that they have had a lot of experience and success in it. However, CLIL learners have been developing reading skills for exactly as many years as in their L1 and have had more experience in mastering it than the control group in both languages. This might be the reason why CLIL learners reported on higher level of L2 reading skills than the control group participants. Since reading needs the complex involvement of the executive functions even in one language, it might be assumed, that CLIL learners are more trained in it. Our assumption is in line with the literature (Engle et al, 1999; Linck et al., 2014) claiming that the higher level of working memory functioning a person has, the better ability they have in learning vocabulary, writing, reading, and listening.

Control group participants report on lower level of their L1 writing and reading skills, while they are proud of their accent and listening skills. In terms of L2, they find reading the least and listening and speaking the most problematic skills. Reading as the second most problematic skill in L1 dominates control group learners' everyday activities, while there is not much emphasis on writing skills. Interestingly, there is an overlap between the two groups in terms of the order of the most dominant skills in L2, in which writing has the most dominant role.

Regarding the activities done in the languages there are also differences between the two groups. Since high target language exposure in out-of-school settings can also contribute to and advanced level of language knowledge, the investigation of these circumstances is crucial (Ball et al., 2015; Attard Montalto et al., 2016). CLIL learners frequently chat with their families, learn or chat with their friends in their L1. They rarely play online games, write texts, or use Skype in this language. Not surprisingly, they do not find learning a dominant activity in their L1 (compared to the control group). In case of the L1 CLIL learners most dominant activities are chatting with friends and family, watching TV and films and info search. Not only CLIL learners' but control group participants' most dominantly pursued L2 activities can be strongly linked to entertainment. These data are not in line with those detailed in the research of Europeans and their Languages (2012), which revealed a more pragmatic language use in case of Hungarians. However, CLIL learners rarely use L2 for family chats, which is in line with the answers given by Hungarians in the same survey. As opposed to other European countries, writing, or contacting people do not feature CLIL learners' dominant activities, mainly because in the school context they might have enough opportunities to express their opinion in the L2. The number of dominant activities in L2 is smaller in case of the CLIL learners, compared to that of the control group participants. Surprisingly, they do not mention learning among their most dominant L2 activities, even though it was found as one of the most dominant activities done in the L2 by Hungarians according to the Europeans and their Languages (2012). We can conclude from the results on CLIL learners' language preferences that they consider their L2 as means of communication instead of a subject to learn.

In terms of control group participants, it can be stated that leisure activities and learning are nearly equally important activities in their L1. Keeping contact with other people is dominated by their L1, however they also practice it in their L2. They use L2 for more varied and productive activities (e.g., writing texts) mainly because they do not have many opportunities for natural language use in instructed settings. Since there are no comparable results on L2 skills for mainstream language learners, only CLIL learners' results on L2 skills could have been compared to those of the national target language test. Overall, it can be concluded that these results on L1 and L2 use are in line with the Complementary Principle (Grosjean, 2016).

4.2. Test d2-R

Test d2-R is a general performance test that measures the ability to suppress interference. During testing top-down (goal-driven) attentional control is engaged. Recent research on bilingualism has assumed that selective attention is engaged while constantly monitoring the context, switching between the languages, and inhibiting irrelevant information (Friesen et al., 2015; Bialystok, 2017). This back-and-forth monitoring and switching process is assumed to result in enhanced domain-general attention, better inhibitory control, respectively.

The reason of selection of Test d2-R in this study was to apply a language-neutral test on selective attention that does not contain an explicit conflict to be resolved. Test d2-R is a visual, language-neutral test measuring selective attention, concentration, and tolerance for monotony. In this study, no significant difference was found between the experimental (CLIL) and the control group in terms of selective attention. For this reason, our hypothesis assuming CLIL learners' significantly higher performance in the test of selective attention in comparison to their peers learning according to a general curriculum must be rejected.

Regarding participants' performance as compared to the age group, it can be concluded that CLIL learners' concentration, accuracy and work pace fall into the average category. Control group participants' performance values are like those of the CLIL group: their results on concentration, accuracy and work pace also fall into the average category. While these findings are inconsistent with those of Bialystok and colleagues' (2010) and Friesen and colleagues' (2015), they are in line with Ratiu and colleagues' (2017) Paap and colleagues' (2018) and Lehtonen and colleague's (2018) who found no significant differences between groups of monolinguals in tests measuring selective attention in the conjunction-search task.

Given that both CLIL and control group participants' results reflect only average level of domain-general attention, the following three possible explanations might be given: a) the extent of L2 experience or proficiency does not influence learners' domain-general attention b) Test d2-R does not capture the cognitive ability in which CLIL learners might outperform the control group, or c) the workload was not high enough for the CLIL learners in this task. Diamond (2013) posited that working memory and selective attention strongly rely on one another. If workload had been higher for the CLIL learners, the number of errors would have increased, indicating lower working memory functioning (Shao et al., 2014). This means that CLIL learners did not have to make extra effort to complete the task. The negligible number of perseverations and repetitions (as signs of improper inhibition) also confirm that working memory was not over-challenged since inhibitory control functioned properly (that is no slowdown in processing was found) (Costa, 2009; Shao et al, 2014; Kousaie et al., 2014; Howard et al., 2014; Diamond, 2016). Overall, I assume that workload in the Test d2-R was not high enough because the test was not verbal in nature. In this sense, my assumption is also in line with Engle and colleagues' (1999) claim that working memory capacity can be differently involved in tasks that cannot be implemented via automatic processing. In other words, if attentional control is strongly related to working memory functioning, their higher involvement might be required in case of higher task load (Calvo et al, 2016).

4.3. Phonemic fluency tests

Fluency tests are usually applied to provide information about the size of the mental lexicon and the cognitive mechanism underlying word retrieval. In the following sections the quantitative analysis of the fluency tests is detailed. It must be noted that up to date no studies have been found applying the same scoring and analysing methods as applied in this study. Most of the studies on evaluating phonemic fluency performance rely on the total number of generated words. However, this value has been questioned to provide quality information about cognitive processing (Troyer, 1997; Kousaie et al., 2014).

In general, we can conclude that CLIL learners did not produce fewer words in terms of the total number of generated words for both test conditions than the control group. The difference between their performances is negligible. These results are in line with the literature assuming the more efficient recruitment of executive functions in the letter fluency task. Given that CLIL learners had a larger vocabulary, they probably had to exclude more words starting with the same letter or being semantically related than the control group learners (Troyer et al., 1997; Luo et al., 2010). These results are partly in line with the normative data produced by Tánczos (2014): both groups performed more words for all the Hungarian test conditions than participants in the normative sample.

Both CLIL and control group learners produced significantly more words in L1 than in L2. Control group participants generated more words in L1 than the CLIL group, but the CLIL group participants were more productive in L2. Considering the total number of words written in L1 and L2, we can conclude that the CLIL group's performance for all letters in the English test (F, A, S) was significantly better compared to the control group's. However, in the Hungarian test (K, T, A) the control group participants produced significantly better result only for letter T. In case of letter K and A no significant differences have been detected, that is no group difference was found. In the light of these data, we can conclude that our hypothesis assuming CLIL learners' significantly better results in the total number of generated words in the English fluency tests is accepted. In contrast, our hypothesis assuming CLIL learners' significantly better results in the total number of generated words in the Hungarian fluency test must be rejected. Since CLIL learners produced significantly more words and not more perseverations than the control group learners in the English phonemic fluency tests, in line with Amunt and colleagues' claim (2020), we can conclude that their working memory performance during the completion task was better.

Regarding the total number of generated words in the L2 phonemic fluency test by bilinguals, our results are in line with previous outcomes gained in various studies (Luo et al, 2010; Marsh et al., 2019; Zeng et al., 2019; Escobar, 2018; Shishkin & Ecke, 2018). Most of the authors claim that better results favouring bilinguals in the phonemic fluency tests are due to the lexical competition between their languages, from which the intrusions from both langauges need to be inhibited. Bialystok and colleagues (2008) posit that compensating for their smaller vocabularies as compared to monolinguals, bilinguals might recruit additional cognitive resources (search strategies) and that makes them able to generate as many words as monolinguals in the phonemic fluency tests. Regarding the strict language-related output requirements of the CLIL programme in Hungary, in case of these learners I would exclude this possibility.

However, I agree with Marsh and colleagues (2019), who claimed that bilinguals' better outcomes might be supported by strategy use like switching and clustering. In the related literature switching is found to be underpinned by general executive functioning (Delgado-Álvarez et al, 2021; Messer et al., 2018) or caused by the lack of clustering ability (mental flexibility) (Abwender et al, 2001.) Clustering, however, is assumed to be underpinned by mental flexibility (Amunts et al, 2020) and provides information about the size of the mental lexicon and the ability to access it (Troyer et al., 1998; Linck et al., 2014; Delgado-Álvarez et al., 2021). To reveal whether learners' strategy use might have contributed to our outcomes, additional markers like the different types of switches and clusters have been calculated. It must be noted that, to the best of our knowledge, no study has been conducted to date aiming an in-depth analysis of strategy use in case of primary Hungarian CLIL (and/or bilingual) and L2 learners. Although, CLIL learners are often reported to perform outstandingly well in tasks requiring switching like in the Dimensional Card Sorting activity or the Simon task (Bialystok et al, 2012), they are not language-related tasks.

In this study, we regarded switching as the indicator of mental flexibility. To have a deeper analysis, two types of switches have been distinguished in the study: hard switches (between single words or a single word and a cluster) and cluster switches (between two clusters). In our study, results revealed significant differences between the two groups in terms of the hard switches for the English phonemic fluency test (FAS), but not for the Hungarian (KTA) fluency test. Regarding cluster switches, significant difference was found in the English phonemic fluency test favouring CLIL learners, therefore our hypothesis is accepted. Since CLIL learners produced significantly more words in total in the English phonemic fluency test and made significantly more switches, they must have recruited more cognitive resources. The reason for this cannot be that they do not have comparable vocabulary, since the strict output requirements of CLIL programmes declared in the corresponding regulations guarantee their vocabulary development. Conversely, they must have nearly similar vocabularies due to the parallel learning of both language and content (Mehisto, 2009; Coyle et al., 2012; Ball et al, 2015; Borowiak, 2019).

Regarding the Hungarian fluency test, however, no significant difference was observed. In our view, the higher number of switches produced by the CLIL group in the English test refers to a greater degree of intentionality in strategy use. Although the fact that

CLIL learners did not have worse results in the Hungarian tests in any variables compared to the control group learners, also confirms that lexical retrieval is not slower or more effortful for them in their L1. Since CLIL learners' supremacy can only be detected in the English fluency tests, we might assume that these results are L2 or L2-context-related. This assumption is supported by CLIL methodology that promotes the application of a vast array of teaching and learning techniques (Mehisto, 2009; Coyle et al., 2012; Ball et al, 2015; Borowiak, 2019). Given that the ability of shifting is highly supported by self-control which is linked to either a personal characteristic or a changeable state (Janka, 2017), we can assume that CLIL learners must have had a higher engagement in the English fluency tests in the context of L2. Since learning in CLIL is a qualitatively different experience we also assume that these learners are used to exhibiting more engagement during learning (Kovács, 2018; Attard Montalto, 2016; Borowiak, 2019). In this sense our claim is in line with Green and Abutalebi's (2007) in terms of importance of interactional contexts of language use in bilinguals' adaptability to various cognitive demands. Overall, in the light of these data we conclude two things: a) CLIL learners had to recruit additional cognitive resources (mental flexibility) to accomplish the English fluency tests and b) the extensive use of L2 does not hinder lexical retrieval from L1 in case of CLIL learners. Our claim is partially in line with that of Bialystok's (2008).

To find a more robust confirmation to this assumption, participants' clustering abilities are also investigated. In this study, we regarded clustering as the ability to use and access the word store (Troyer et al, 1997). According to the literature, working memory has a huge role in the access and retrieval of information from the long-term memory (Meltzer, 2014; Miyake et al., 2000; Diamond, 2013, 2016). In the light of the literature, we can state that these CLIL learners have a better access to the mental lexicon in their L2 and there is no difference between the two groups in terms of L1. Consequently, the involvement of working memory function was higher in case of CLIL learners in the English phonemic fluency tests. Our results also revealed that CLIL learners produced significantly more task-discrepant (semantic) and task-consistent (phonological) clusters compared to the control group learners in L2, therefore our hypothesis is accepted. In our study, we followed Abwender's claim (2001) and considered the creation of task-discrepant clusters as the real signs of intentional strategy use. The fact that CLIL learners produced twice as many task-discrepant clusters compared to the control group learners suggests that they recruited additional cognitive resources. It is widely held that that under the phonemic fluency condition, participants generate more task-consistent solutions. However, our findings show a different pattern. The possible explanation is that meaning is in the focus of CLIL methodology. To gain information about learners' organization of the semantic memory, we investigated the sizes of the generated clusters. Our results show that CLIL group participants have created significantly larger clusters for each letter condition in the English phonemic fluency test than the control group participants, thus our hypothesis is accepted. Regarding cluster sizes in the English phonemic fluency tests, our results are in line with those of Patra and colleagues (2020). This finding also confirms our assumption that CLIL learners focus on meaning while using L2. As CLIL is considered as constructivist pedagogy that focuses on meaning-making during the learning process, it promotes linking concepts. For this reason, it is not unusual when learners of these programmes apply methods (linking words) in language-based tests accordingly. Surprisingly though, no significant differences in cluster sizes were found in the Hungarian phonemic fluency test, therefore our hypothesis is rejected. Regarding the number of errors, repetitions and perseverations no significant differences have been found between the two groups in terms of either test condition (Hungarian and English). If we regard the appearance of perseverations as clear signs of deficit in inhibition (Shao et al, 2014), we might conclude that CLIL group participants are able to inhibit intrusions just as well as the control group participants. However, these results are not in line with the normative data reported on the Hungarian fluency test outcomes by Tánczos (2014). Both groups in our study outperformed the normative sample in all letter conditions.

At this point the possible other factors that might have had impact on the test outcomes are under scrutiny. Zeng and colleagues (2019) propose that bilinguals' positive linguistic and non-linguistic test outcomes might be caused by ongoing developmental changes. Given that the development of executive functions that underlie verbal fluency and selective attention test outcomes last longer than adolescence (Malute et al., 2004; Zelazo et al., 2016; Diamond, 2013; Diamond, 2016), and we consider that control group participants are also exposed to L2 both at school and at home, we expect closer results for both groups in terms of selective attention and the verbal fluency tests. Even if learners' cognitive development follows a different pattern, our results suggest that developmental changes might not have caused these test outcomes.

At the same time, other researchers highlight the roles of both the vocabulary size and the involvement of the executive functions in the test outcomes (Luo et al., 2010; Escobar, 2018; Friesen et al., 2021). Since CLIL learners must have a certain lexical overlap between their languages because of parallel academic vocabulary learning, we might assume that their vocabulary in L1 and L2 must be more balanced. If a greater vocabulary size is, in general, paired with better executive functioning, CLIL participants in our study should have performed equally in both fluency tests. Since no significant difference was found between CLIL and control group participants in terms of the Hungarian test, we might assume that CLIL learners had to recruit additional cognitive resources to complete the English fluency tests.

Overall, we can conclude that if the task load is greater for Hungarian CLIL learners in tasks requiring the involvement of executive functions, they outperform their matched peers. These finding are in line with Linck and colleagues' (2014), Calvo and colleagues' (2016); Antón and colleagues' (2019). We also share Antón and colleagues' (2019) claim that bilingualism might have impact only on specific cognitive constructs.

4.3.1. Word types

Regarding word types, 50-60% of the generated words in these test types are nouns and 15-20% of them are verbs. Numbers, modals and adverbs are quite rare (Gósy, 2005; Navracsics, 2007b; Lukács et al., 2014). Our results confirm this only in case of the English phonemic fluency test, where 57.07% of the total FAS words created by the CLIL group are nouns. In case of the control group this number is 50%. It can also be concluded that CLIL group participants produced significantly more abstract nouns and concrete nouns in the English phonemic fluency test in comparison with the control group participants. This growth in the number of abstract nouns is in line with our expectations and the related literature: as abstract knowledge develops, so does the underlying vocabulary (Cummins, 2000; Lukács et al., 2014). However, the results of the Hungarian phonemic fluency test (KTA) in terms of the proportion of nouns differ from our expectations: 76.26% of the total KTA words created by the CLIL group are nouns. In case of the control group this number is 78.19%. Although the control group participants created more abstract nouns in the Hungarian phonemic fluency test, no significant difference was found between the two groups. Regarding concrete nouns in the Hungarian test, control group participants created significantly more words. If we accept that processing and acquisition of abstract concepts are more difficult and therefore occur at a later age as compared to concrete nouns, we might conclude that CLIL learners' retrieval ability in L2 is more developed compared to the control group participants'. Furthermore, the fact that no significant difference was found between the two groups in terms of abstract nouns generated in the Hungarian phonemic fluency test, also confirms our assumption that retrieval is not more effortful for the CLIL learners.

Considering the proportion of generated verbs, the results are even more striking. In the English phonemic fluency test, the results are in accordance with our expectations. 18.20% of the total FAS words created by the CLIL group are verbs. In case of the control group this number is 21.41%. However, results of both groups in the Hungarian phonemic fluency test report on weaker results than expected. 10.53% of the words created by the CLIL group participants in the KTA test are verbs. This number is 11.44% in case of the control group. These results are partially in line with the literature (Gósy, 2005; Navracsics, 2007b; Lukács et al., 2014).

Regarding the total number of adjectives generated in the English phonemic fluency test, we can conclude that the CLIL group participants created significantly more of them than the control group participants. However, the distribution level of adjectives is lower in both language tests for both groups in comparison with that of other word types. In the English phonemic fluency test, the proportion of adjectives is 15.16% for the CLIL and 12.36% for the control group. Regarding the Hungarian phonemic fluency test, the results are: 6.93% for the CLIL and 5.95% for the control group.

4.3.2. Unique words, genre and word frequency

In order to gain a deeper analysis, we applied an unusual approach of investigating the two groups' vocabulary sturcture and size. Namely, the net number of vocabulary items specific to one of the groups was calculated. Our results revealed that the CLIL learners produced about twice as many unique words for each letter than the control group learners in the English (FAS) phonemic fluency test. The distribution of the different genres of these unique words were also imbalanced between the two groups. The CLIL group created more unique words for each genre and letter from which the proportions of vocabulary, related to fiction, magazine and academic language were outstanding. These data are in line with the dominant acitivities (reading and watching films) detailed in the LEAP-Q questionnaire. The slightest differences between the two groups in terms of the proportion of genres were detected in case of vocabulary related to spoken language and newspaper. Surprisingly, these data are not in line with results of the LEAP-Q questionnaire. Regarding frequency of unique words, we conlcuded that CLIL learners produced more frequent and rare unique words in comparison with the control group. The higher proportion of rare words, however, might also confirm that CLIL learners' vocabulary size is not lower than that of the control participants, who also produced more rare words than frequent ones. Given that rare words are mainly subject-specific, they are highly linked to thinking skills. These results suggest the influential role of frequency in word retrieval (Gósy, 2005; Larsen-Freeman, 2012; Ball et al., 2015 Ellis, 2015).

Regarding the unique words in the Hungarian phonemic fluency test (KTA), we concluded that the differences between the two groups for each letter can be expressed only in a few percentages, that is the data are more balanced. Regarding the genres, the picture was more mixed: CLIL learners created more unique words related to press and official language use, while control participants produced more of them related to literature, spoken language and academic language. These data are in line with the results of the LEAP-Q questionnaire in case of the control group. However, the dominance of lexical items related to press and official language in case of the CLIL group is at least surprising.

4.4. Structured interviews

4.4.1. Aim of learning

In general, we can conclude that both 'high' groups considered English as a lingua franca, e.g., a common language that reflects a pragmatic view of L2 learning. CLIL high learners provided more and more varied answers than their control group peers. While control group learners focused rather on the four L2 skills, CLIL learners also highlighted the importance of pronunciation in L2 learning, which is in line with Navracsics' claim (2008). They also emphasized the importance of self-confidence described as a focal point in CLIL methodology (Mehisto et al., 2009; Borowiak, 2019) and a crucial 21st century competence (Attard Montalto et al., 2016; McGuinness, 2018). Control 'high' group participants approached the reasons of L2 learning from three perspectives: a social aspect, a functional aspect, and a skill. However, CLIL ‘high’ group participants added two more aspects to the previous three: an emotional aspect and a personal goal referring to a wider spectrum of their thinking. Since some of the CLIL learners found L2 interesting and motivating, a more positive attitude emerges from their responses. This is one of the aims of a competence-based curriculum (Marope et al., 2017). In general, it can be stated that the ‘low’ group’s reflections are not as complex as those of the ‘high’ group. Control ‘low’ group participants’ thoughts are listed around four main aspects: a social, an emotional, a functional and a skill. In contrast, CLIL ‘low’ group participants’ reflections revolve only around three main nodes: skills, functionality, and personal goals. Adding ‘personal goals’ to the list is only a CLIL group characteristic. CLIL ‘low’ group learners’ answers covered all the main L2 skills, while their peers focused more on communication.

4.4.2. Characteristics of a good language teacher

Complexity and diversity of reflections can be clearly captured in the CLIL ‘high’ learners’ answers compared to those of the control group. Their visions and expectations of a good language teacher’s characteristics confirm the existence of different teaching methods in the lessons. Moreover, the ability to notice these might refer to a high L2-related awareness. While control ‘high’ participants approached the issue along two clear dimensions (teacher’s personality and class managing techniques), CLIL ‘high’ group participants listed teaching methods also. The mentioning of usage-based L2 learning, the various tasks and practices and the learning strategy are all in accordance with the literature on CLIL methodology (Mehisto et al., 2009; Pérez-Cañado (2012); Attard Montalto et al., 2016; Borowiak, 2019) and competence-based teaching. The different learning environment results in a different L2 atmosphere (Kovács, 2018). The flow of ideas does not only confirm the CLIL ‘high’ learners’ ability of breaking away from reality, fast retrieval, critical thinking but shifting among their ideas as well. ‘Low’ group learners, as opposed to ‘high’ group learners, only set expectations regarding the teachers’ responsibilities and they seem to be passive recipients in the learning situation. Given that, both control groups (high and low) approached the issue from the same perspectives and provided only a limited number of ideas, our assumption is that these learners cannot mentally break away from reality.

4.4.3. L2 lessons or lessons in L2

In general, we can conclude that CLIL learners approach the question of a quality L2 lessons or lessons in L2 more positively. They report on a good classroom atmosphere in which they do various activities. They list eight different ideas that might characterize a good L2 lesson. These can be subdivided into three groups along the following aspects: pace of the lesson, atmosphere, and task-types. This methodologically enriched environment is in line with CLIL principles (Mehisto, et al., 2009; Attard Montalto et al., 2016; Borowiak, 2019). In contrast, control 'high' group learners report on monotonous lessons in which community seems to be the main motivating factor. In case of the 'low' groups, there is a high consistency among the opinions and the less positive attitude towards the L2 lessons or lessons in L2. While CLIL group participants report on lessons a bit more positively, control group participants seem to be less enthusiastic. In general, we can conclude that both 'low' groups listed the same and the same number of aspects (referring to the task-types: monotonous and varied).

4.4.4. Activities

In general, we can conclude that CLIL group learners were able to list more L2-related activities than control group participants. Preferred activities listed by the CLIL 'high' group cover three of the main language skills (listening, reading, and writing) and are in line with the results of the LEAP-Q questionnaire. In contrast, control 'high' group participants listed one preferred unusual task that is related to enjoyment. Regarding the disliked activities, the same tendency can be observed. CLIL 'high' group learners specified more task types that can be grouped along two aspects: unusual task-type and skill-related activities, two of which (speaking and writing) were identified as their weaknesses in the LEAP-Q questionnaire. Conversely, control 'high' participants focused only on skills (listening and reading-translation) while listing disliked activities among which listening comprehension is identified as their weakest skill in L2 in the LEAP-Q questionnaire. Control group learners' comfort-safety attitude towards L2 can also be detected in their answers. The small number of reflections provided by the control group suggests that these learners are unaware of the activities they must do in the lessons; therefore they are unable to name any of them explicitly or shift perspectives to express their wishes. CLIL 'low' group participants listed more preferred and disliked activities than their peers, although there is no difference between the two groups in terms of the number of aspects (skill-based and unusual) along which these activities were characterized. Regarding disliked activities control group participants do not seem to be able to think out of the 'old-school' teaching methods again, while CLIL participants listed more specific activities.

4.4.5. Development

CLIL 'high' learners' tendentiously more complex and diverse approaches can also be detected on the issue of their L2 development about which they are much more positive than the control group learners. CLIL learners approached this topic from two main perspectives: development of L2 skills and that of personality. Considering this question, there is only a partial overlap with the answers given by them in the LEAP-Q questionnaire, in which reading, and listening are defined as their strengths and speaking is identified as one of their weaknesses. However, it does not seem a real contradiction, since the issue is not the level of speaking skills, but a supporting feature beyond it (being more confident). CLIL group learners seem to be more conscious about their L2 development as compared to the control group participants. In terms of L2 development control 'low' group participants provided more complex and diverse reflections compared to the CLIL 'low' group participants. These reflections can be grouped into two categories: that of L2 skills and personal characteristics. CLIL group participants, however, listed only two L2-components that can be grouped into one category.

4.4.6. Competences

In general, we can conclude that CLIL 'high' participants produced ideas that can be grouped in three main categories: skills, cognitive ability, and personal L2 strategy. In contrast, control group participants listed words are related to skills and a cognitive ability. CLIL group participants provided more and more diverse reflections especially in case of strengths, although their answers seem to be strongly related to the B2 level exam they were preparing for. In terms of weakness they were also more critical: all of them considered grammar as a crucial point in their L2 knowledge. Assuming that their teachers apply CLIL methodology, grammar might not be a focal point in the learning process, since meaning-attribution is the essence of knowledge construction in this programme (Mehisto et al., 2009; Borowiak, 2019; Goris, 2019). In terms of weaknesses, control group participants seem to be unable to approach the question from a different perspective: grammar and translation are dominant in their reflections. In 'low' groups, there is also a tendency for a more balanced number and type of reflections. In general, we can conclude that both groups listed items that belong to the same class (L2 skills) and could not change their viewpoint or be more specific.

4.4.7. Do differently

Furthermore, we can conclude that while CLIL 'high' learners' ideas on what they would do differently in the L2 classroom can be grouped into two categories, control 'high' learners provided words that all belong to the same category. The list of these specific ideas confirms our assumption that CLIL 'high' learners have more diverse mindsets, and they are also able to change their focus with ease.

'Low' group learners, however, provided more similar answers. In general, they listed less sub-codes than the 'high' groups and they seem to settle in a comfort-safety position and be entertained by the teacher in the lesson without taking an active part in it. Participants in both groups listed ideas that can be divided into two categories: teacher's inner feature (e.g. not strict) and learner-centred activity types (e.g. learning through games).

4.4.8. Differences in wording and unanswered questions

CLIL learners' statements were formulated in a stylistically expected manner, some control group learners used word phrases that did not match the style and the topic articulated in the question. CLIL learners generally express their opinions in longer and structurally more complex sentences and tend to use more adjectives than their control counterparts.

If the number of unanswered questions is regarded, the lack of goal-oriented behaviour can be detected. While both 'high' group participants answered all the questions, some 'low' group participants missed a few of them.

CHAPTER 5. Conclusions and future directions

Today an individual's adaptability of thoughts and behavioral patterns to the dynamically changing contexts is considered a high potential at the labour market. This adaptability might cover many different competences from the burst of ideas to efficient L2 communication that can be grounded in primary education. Although, the educational system in Hungary is currently in a state of transition, most teachers are not trained or prepared for the introduction of those methods and practices that might support learners' integration into the work environment of the future.

The number of primary educational programmes that favour practices other than the ones applied in the traditional 'drill-and-skill' education is increasing. Due to the different educational and pedagogical goals set by CLIL programmes, these teaching techniques and practices have been present for a while. In such programmes the boundaries between different school subjects are blurred, so is the relationship between teachers and learners. The different L2 teaching context can result in learners' qualitatively different levels of knowledge, learning paths and mental sets.

Based on or results we can state that extensive (daily) experience with an L2 in and out of instructed conditions subserves the emergence of a vast array of positive outcomes. Firstly, language learners become real users of the L2, therefore they not only get engaged in formal or informal situations with ease, but quickly adjust to the changing tasks or language contexts. The higher expectations placed on learners by CLIL can also be assumed from their self-assessed level of own language skills. Altough, learners' rigorous evaluation on their own language skills do not discourage them from using the L2 for entertainment on online and offline platforms where it is the medium of communication. This way, their attitude towards L2 use is considered positive. Secondly, CLIL programmes are often referred to as dual-language programmes. As this name implies, the parallel use of both languages in content and language classes is a natural feature of the programme which is aimed at supporting the process of conceptual knowledge construction. Consequently, there must be a great overlap between the lexical items of the two languages at least in terms of academic terminology. In this study, results of phonemic fluency tests support this assumption. Moreover, CLIL learners seem to apply more cognitive strategies to retrieve lexical items, which can be seen as the indicator of mental flexibility. Thirdly, CLIL learners are accustomed to being actively involved in task types that require competence in problem-solving, concluding and cooperation even in the L2. As a result of meaningful practice, they also participate in these activities with greater enthusiasm. CLIL learners consider innovative teaching practices as norms and desires at the same time and they are able to recognize their emergence during the learning process. Furthermore, they are not clueless if asked to express their opinions on various topics and tend to approach the issues from different aspects that can be seen as an indicator of mental flexibility.

However, CLIL participants' outstanding results might be linked to their inherently higher cognitive functioning or the special teaching practices applied in CLIL programmes. In the following sections our views on these issues are explained. In studies focusing on bilingual participants' cognitive functions, there is always 'the chicken or the egg' casuality dilemma, that is, the reason why research participants are able to perform better in tests measuring cognitive abilities is due to the originally better operations of their executive functions. In terms of our study, this assumption can be rejected for some reasons. Firstly, participants have not gone through a selection process in order to be enrolled at their primary schools, which fact is the warranty on the presence of mixed abilitiy learners in both types of educational programmes. Secondly, learners participated in the research anonymously. Finally, data on participants' socioeconomic status revealed that, there were no remarkable differences between the groups in terms of their family backgrounds that might have had impact on the research outcomes. Moreover, even the parameters of selected learners were very similar regarding their test results and socioeconomic status. It is important to note that in this study we sought to avoid the application of subjective factors related to socioeconomic status (such as the number of books owned by the family), and use objective ones (such as parents' highest level of education, marital status or employment).

Another dilemma that should be considered under the scope of bilingualism is whether CLIL learners' higher performance is the direct consequence of CLIL methodology and not that of extensive L2 use. Approaching the issue in terms of applied teaching practices we can state that although these programmes work with well-defined methodology it is not guaranteed that all CLIL teachers continue this practice on a daily basis, as it would require excessive preparation, which is difficult to be implemented.

Due to the high number of lessons in the CLIL programmes and learners' advanced level of L2 proficiency, there are more opportunities for integrating unusual topics and task types in the lessons. Although, statutory requirements on language and content outcomes leave the teachers only a little room for maneuver. In addition, a teacher's personality and teaching style can also have impact on the efficacy of the lesson. Furthermore, those conditions and tools (computers, smart boards, whiteboards, internet-access, printers) that could make a lesson more interactive and variable are often limited. For this reason, their intensive application is not typical in an average CLIL class. Furthermore, if CLIL methodology had indeed had such an impact on learners' mindsets, interview results would have been even more complex and specific in case of the CLIL 'low' group.

Based on our results, we can generally conclude that, CLIL 'high' learners were able to express their views on L2-related issues in a qualitatively different way than learners in the comparable control groups. Given that, selected participants were matched for their socioeconomic status, level of concentration, verbal abilities and executive functioning, and CLIL as an L2 approach might not have had an exlusive role in learners' mindsets, we can assume that extensive L2 experience coupled with CLIL methodology might have caused different research outcomes in case of the CLIL learners.

5.1. Future directions

Finally, we can conclude that the applied methods were in line with the aims of the research and our quantitative results converged with the qualitative findings. However, an in-depth analysis would reveal those hidden factors that might have had influence on the research outcomes. As a continuation of the research, unstructured interviews with CLIL teachers of the participating schools would provide additional information related to their current methodological repertioire. This way CLIL theory and practice could be compared.

5.2. Limitations of the research carried out

Despite the methodologically well-thought-out research design, I had to face some unforseen limitation that I am discussing in the following section.

The most determinig difficulty I had to face was the number of participants. Originally, I intended to involve at least two hundred learners, half of whom would have represented the experimental group. When I contacted school principals and asked for their consent for the research, some of them proved to be open and proactive during preparation and implementation, while others did not even respond to my inquiry. Since the planned phase of data collection ended in December 2019, shortly before the emergence of the pandemic, I no longer had the opportunity to involе more institutions in the research.

5.3. Contributions to theory and practice

Until the previous decade the determining role of executive functions in a child's academic success was underestimated among researchers and practicing teachers as well. By now, it is well evidenced that their high-level operation is not only the basis of attentional, thinking and problem-solving processes, but that of well-sustained behavioural patterns, flexible adoption of new perspectives and adjustment to rules as well. These are those core skills on which learning outcomes highly depend and their state in early childhood is predictive regarding future executive functioning. Children who are coping with learning difficulties have deficits in their executive functions, which have a negative impact on academic achievement and also on social behaviour. Students with learning disabilities not only suffer from the lack of effective learning strategies or self-regulation, they are also highly challenged when a task needs planning, organizing and sorting out information, shifting strategies, flexible thinking or metacognition.

Due to the malleability of the prefrontal cortex, any kind of direct or indirect training, can have positive or negative influence on brain development and that of the executive functions. Educational practice and school context have a huge responsibility in their improvement because academic learning not only improves executive functions and vice versa, but they are proved to be protective against risks of children's disadvantageous life conditions as well. Consequently, instead of drill-and-skill education, play-based methods should be integrated in the curriculum since they meet with children's developmental needs and can promote their academic skills.

In this study, I aimed to reveal whether intensive exposure to L2 in instructed settings might influence learners' lexical retrieval in L1 and L2. I also intended to investigate whether CLIL learners, as a result of the intensive exposure to L2, can be characterized by intentional strategy use during lexical retrieval in phonemic fluency tests. My results revealed that intensive L2 use does not have a negative influence on CLIL learners' lexical retrieval but promotes strategy use not only in word retrieval but in an interview situation as well. Findings of this study have confirmed my assumption that outcomes of usage-based teaching and learning methods might go far beyond a B2-level exam.

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3), 242-275.

Abwender, D., Swan, J., Bowerman, J. & Connolly, S. (2001). Qualitative analysis of verbal fluency output: review and comparison of several scoring methods. *Assessment*, *8*(3).

Amunts, J., Camilleri, J., Eickhoff, S., Heim, S., & Weis, S. (2020). Executive functions predict verbal fluency scores in healthy participants. *Scientific Reports*, *10*(1).

Antón, E., Carreiras, M., & Duñabeitia, J. (2019). The impact of bilingualism on executive functions and working memory in young adults. *PLOS ONE*, *14*(2), e0206770.

Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In: Bower, G. (Ed.). *Recent Advances in Learning and Motivation.* Vol. VIII. Academic Press, New York, 47-90.

Ball, P., Kelly, K. & Clegg, J. (2015) *Putting CLIL into Practice*. Oxford: Oxford University Press.

Benton, A. L., Hamsher, de, S. K., & Sivan, A. B. (1983). *Multilingual aphasia examination* (2nd ed.). Iowa City, IA: AJA Associates.

Bialystok, E., Craik, F., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends In Cognitive Sciences*, *16*(4), 240-250.

Bialystok, E. & Poarch, G. (2014). Language experience changes language and cognitive ability. *Zeitschrift Für Erziehungswissenschaft*, *17*(3), 433-446.

Bialystok, E. (2009): Consequences of bilingualism for cognitive development. In Kroll, J. F. & Groot, A. (Eds.) *Handbook of bilingualism* (pp. 417-352). Oxford, Oxford University Press.

Bialystok, E. (2010). Global-local and Trail-making Tasks by Monolingual and Bilingual Children: Beyond Inhibition. *Developmental Psychology*, *46*(1) 93-105.

Bialystok, E. (2017).: How minds accommodate experience. *Psychological Bulletin*, *143*(3), 233-262.

Biesbroek, J., van Zandvoort, M., Kappelle, L., Velthuis, B., Biessels, G., & Postma, A. (2015). Shared and distinct anatomical correlates of semantic and phonemic fluency revealed by lesion-symptom mapping in patients with ischemic stroke. *Brain Structure And Function*, *221*(4), 2123-2134.

Borowiak, A. (2019). CLIL Education. In: Lewandowska-Tomaszczyk B. (Eds) *Contacts and Contrasts in Educational Contexts and Translation. Second Language Learning and Teaching*. (pp. 73-83). Springer, Cham.

Braem, S., & Egner, T. (2018). Getting a grip on cognitive flexibility. *Current directions in psychological science*, *27*(6), 470–476.

Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2010). *Test d2 - Revision*. Göttingen: Hogrefe.

Buttelmann, F. & Karbach, J. (2017). Development and plasticity of cognitive flexibility in early and middle childhood. *Frontiers In Psychology*, *8*.

Calvo, N., Ibáñez, A., & García, A. (2016). The impact of bilingualism on working memory: a null effect on the whole may not be so on the parts. *Frontiers In Psychology*, *7*.

Carone, D. (2007). E. Strauss, E. M. S. Sherman, & O. Spreen,A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. *Applied Neuropsychology*, *14*(1), 62-63.

Cook, V. & Singleton, D. (2014). *Key topics in second language acquisition*. Bristol: Multilingual Matters.

Costa, A. (2009): Lexical Access in Bilingual Production. In Kroll, J. F., & Groot, A. (Eds.) *Handbook of bilingualism*. p. 308-325. Oxford: Oxford University Press.

Cowan, N. (1999). *An Embedded-Processes Model of working memory.* In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (p. 62–101). Cambridge University Press.

Coyle, D., Hood, P., & Marsh, D. (2012). *CLIL*. Cambridge: Cambridge University Press.

Creswell, J. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research*. New Jersey: Pearson Education, Upper Saddle River.

Cummins, J. (2000) Language, power and pedagogy: Bilingual children in the crossfire. *Bilingual Education and Bilingualism.* Clevedon, UK: *Multilingual Matters*.

Csépe, V. – Győri, M. – Ragó, A. (2007). Viselkedéskontroll és megismerés: a végrehajtó működések. In *Általános Pszichológia 3.* (pp. 322-335). Budapest: Osiris.

Csépes, I. (2019). Nyelvtudásmérési és -értékelési műveltség: a magyarországi angoltanár (tovább)képzés aktuális kihívásai. *Pedagógusképzés*, *18*(3-4), 77-104.

Csíkos, Cs. (2009). *Mintavétel a kvantitatív pedagógiai kutatásban*. Budapest: Gondolat Kiadó.

Dajani, D., & Uddin, L. (2017). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends In Neurosciences*, *38*(9), 571-578.

Dallinger, S., Jonkmann, K., Hollm, J., & Fiege, C. (2016). The effect of content and language integrated learning on students' English and history competences – Killing two birds with one stone? *Learning and Instruction*, *41*, 23-31.

Dancsó, T. (2007): Az informatikai kompetencia fejlesztési lehetőségei és az IKT-eszközök alkalmazása az oktatásban. In: Bábosik István és Torgyik Judit (szerk.): *Pedagógusmesterség az Európai Unióban*. Eötvös József Kiadó, Budapest. 67-85.

Daucourt, M., Schatschneider, C., Connor, C., Al Otaiba, S. & Hart, S. (2018). Inhibition, Updating Working Memory, and Shifting Predict Reading Disability Symptoms in a Hybrid Model: Project KIDS. *Frontiers In Psychology*, *9*.

De Boer, M., & Leontjev, D. (2020). Conceptualising Assessment and Learning in the CLIL Context. An Introduction. In M. deBoer & D. Leontjev, *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms* (1st ed., pp. 9-24). Springer.

De Bot, K., Chan H., Lowie, W., Plat, R. & Verspoor, M. (2012). A dynamic perspective on language processing and development. *Dutch Journal of Applied Linguistics 1*(2), 188-218.

De Bot, K., Lowie, W., Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition 10* 7-21.

De Groot, A. & Dukes, P. (2011). *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. New York: Psychology Press.

De Groot, A. (2013): Bilingual Memory. In Grosjean, F., Li, P. (Eds.), *The psycholinguistics of bilingualism.* (pp. 171-186). Wiley-Blackwell Publishing.

De Luca, C. & Leventer, R. (2008). Developmental Trajectories of Executive Functions across the Lifespan. In V. Anderson, R. Jacobs & P. (Eds.) Anderson, *Executive Functions and the Frontal Lobes: A Lifespan Perspective,* Taylor & Francis, New York, 23-56.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22,499–533.

Diamond, A. (2013). Executive Functions. *Annual Review Of Psychology, 64*(1), 135-168.

Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), E*xecutive function in pre-school-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 11–43). Washington, DC: American Psychological Association.

Einhorn, Á. (2015a). *A pedagógiai modernizáció és az idegennyelvtanítás*. Miskolci Egyetemi Kiadó, Miskolc.

Einhorn, Á. (2015b): Pedagógiai kultúraváltás az idegennyelv-tanításban? Vitaindító. *Modern Nyelvoktatás*, 21. évf., 4. sz. 48–58.

Ellis, N. C. (2015). Implicit and explicit learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3-23). Amsterdam: John Benjamins.

Engle, R. (2018). Working Memory and Executive Attention: A Revisit. *Perspectives On Psychological Science*, *13*(2), 190-193.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge University Press.

Escobar Urmeneta, C. (2019). An Introduction to Content and Language Integrated Learning (CLIL) for Teachers and Teacher Educators. *CLIL. Journal Of Innovation And Research In Plurilingual And Pluricultural Education*, *2*(1), 7.

Falus, I. & Kimmel, M. (2003): *A portfólió.* Gondolat Kiadó, Budapest.

Fenesi, B., Sana, F., Kim, J. A., & Shore, D. I. (2015). Reconceptualizing working memory in educational research. *Educational Psychology Review*, *27*(2), 333–351.

Friedman, N.P. & Miyake, A. (2004). The relations among inhibition and interference cognitive functions: A latent variable analysis. *Journal of Experimental Psychology: General*, *133*, 101–135.

Friesen, D., Edwards, K., & Lamoureux, C. (2021). Predictors of verbal fluency performance in monolingual and bilingual children: The interactive role of English receptive vocabulary and fluid intelligence. *Journal Of Communication Disorders, 89 (2021)*, 106074.

Gärtner, A., & Strobel, A. (2021). Individual Differences in Inhibitory Control: A latent Variable Analysis. *Journal of Cognition, 4*(1): 17, pp.1–18.

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*, *18*(6), 675–698.

Gósy M. (2005). *Pszicholongvisztika* Budapest, Osiris.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81.

Green, D., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal Of Cognitive Psychology*, *25*(5), 515-530.

Griffiths, C. (2018). *The strategy factor in successful language learning. The tornado effect.* 2nd edition. Multilingual Matters.

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

Grosjean, F. (2013). Bilingualism: A Short Introduction. In: Grosjean, F., Li, P. (Eds.): *The psycholinguistics of bilingualism*. Hoboken: Wiley-Blackwell/John Wiley & Sons. (pp. 5-25).

Grosjean, F. (2016). The Complementarity Principle and its impact on processing, acquisition, and dominance. In C. Silva-Corvala´n & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 36–65). Cambridge: Cambridge University Press.

Gunnerud, H., ten Braak, D., Reikerås, E., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, *146*(12), 1059-1083.

Győri, M. (2008). Viselkedéskontroll és megismerés: a végrehajtó működések. In Csépe V., Győri M., & Ragó A. (Eds.) *Általános pszichológia 3.*: *Nyelv, tudat, gondolkodás* (pp.321-356). Budapest: Osiris.

Hercz Mária (2007): A értékelés gyakorlata. In: Bábosik István és Torgyik Judit (Eds): *Pedagógusmesterség az Európai Unióban* (pp. 191-241) Eötvös József Kiadó, Budapest.

Howard, K., Anderson P. J. & H G. Taylor (2008). Executive functioning and attention in children bor preterm. In V. Anderson, R. Jacobs & P. Anderson, *Executive Functions and the Frontal Lobes A Lifespan Perspective* (pp. 219-234). New York: Psychology Press.

Howard, S., Johnson, J., & Pascual-Leone, J. (2014). Clarifying inhibitory control: Diversity and development of attentional inhibition. *Cognitive Development*, *31*, 1-21.

Janacsek, K., Tánczos, T., Mészáros, T. & Németh, D. (2009). A munkamemória új magyar nyelvű neuropszichológiai mérőeljárása: A hallási mondatterjedelem teszt (HMT). *Magyar Pszichológiai Szemle*, 64(2), 385–406.

Janka, Z. (2017). A mentális rugalmasság idegtudománya. *Orvosi Hetilap*, *158*(45), 1771-1786.

Jessner, U. (2008). Teaching third languages: Findings, trends and challenges. *Language Teaching*, *41*(1), 15-56.

Jessner, U. (2008). A DST model of multilingualism and the role of metalinguistic awareness. *The Modern Language Journal*, *92*(2), 270-283.

Kalia, V., Wilbourn, M. P. & Ghio, K. (2014) Better early or late? Examining the influence of age of exposure and language proficiency on executive function in early and late bilinguals. *Journal of Cognitive Psychology*, 26 (7), 699-713.

Kaushanskaya, M., Blumenfeld, H., & Marian, V. (2019). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, *23*(5), 945-950.

Kavé, G., Kigel, S. & Kochva, R. (2008). Switching and clustering in verbal fluency tasks throughout childhood. *Journal of Clinical And Experimental Neuropsychology*, *30*(3), 349-359.

Kenneth R. Paap & Zachary I. Greenberg (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, Volume 66 (2), 232-258.

Kousaie, S. (2014). Executive function and bilingualism in young and older adults. *Frontiers In Behavioural Neuroscience*, *8*.

Kovács. J. (2009). *A gyermek és az idegen nyelv*. Eötvös J. Könyvkiadó, Budapest.

Kovács, J. (2018). *Iskola, nyelv, siker*. Eötvös József Könyvkiadó, Budapest.

Kovács, J. (2006). *Magyar-angol kéttannyelvű általános iskolai programok közoktatásunkban*. Eötvös József Könyvkiadó, Budapest.

Kuckartz, U. (2019). Qualitative Text Analysis: A Systematic Approach. In Kaiser, G. & Presmeg, N. (Eds), *Compendium for Early Career Researchers in Mathematics Education* (pp. 181-197.).

Kunschak C. (2020) CEFR, CLIL, LOA, and TBLT – Synergising Goals, Methods and Assessment to Optimise Active Student Learning. In: deBoer M., Leontjev D. (eds) Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms. Springer, Cham.

Larsen-Freeman, D. (2012). On the roles of repetition in language teaching and learning. *Applied Linguistics Review*, *3*(2), 195-210.

Lasagabaster, D. (2008). Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal*, *1,* 30-41.

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological bulletin*, *144*(4), 394–425.

Leite, G., Pires, I., Aragão, L., Paula, A., Gomes, E., & Garcia, D. et al. (2016). Performance of Children in Phonemic and Semantic Verbal Fluency Tasks. *Psico-USF*, *21*(2), 293-304.

Lenneberg, E. (1967). *Biological foundations of language*. Robert E. Krieger, New York.

Lesznyák, M., (1996): Kétnyelvűség és kéttannyelvű oktatás. In: *Magyar Pedagógia*, *96*(3), 217-230.

Linck, J., Osthus, P., Koeth, J. & Bunting, M. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861-883.

Lowie, W. & Verspoor, M. (2019). Individual Differences and the Ergodicity Problem. *Language Learning*, *69*, 184-206.

Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, *14*, 588–595.

Lukács, Á., Pléh, Cs., Kas, B. & Thuma, O. (2014). A szavak mentális reprezentációja és az alaktani feldolgozás. In Pléh, Cs., Lukács, Á. (Eds.). *Pszicholingvisztika 1*. (pp. 167-250), Akadémiai Kiadó, Budapest.

Luo, L., Luk, G. & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, *114*(1), 29-41.

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.

Mehisto, P., Frigols, M., & Marsh, D. (2009). *Uncovering CLIL*. Oxford: Macmillan.

Meltzer, L. (2014): Teaching Executive Functioning Processes: Promoting Metacognition, Strategy Use, and Effort. In: Goldstein, S., Naglieri, J. A. (Eds.) *Handbook of Executive Functioning* (pp. 495-508).

Messer, D., Bernardi, M., Botting, N., Hill, E., Nash, G., Leonard, H. & Henry, L. (2018). An Exploration of the Factor Structure of Executive Functioning in Children. *Frontiers in Psychology, 9*.

Mészáros, A., Kónya, A. & Kas, B. (2011). A verbális fluenciatesztek felvételének és értékelésének módszertana. *Alkalmazott Pszichológia 2011*(2), 53-76.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14.

Miyake, A., Friedman, N., Emerson, M., Witzki, A., Howerter, A., & Wager, T. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49-100.

Nádasi, M. (2000). A kikérdezés. In Falus (Ed.), *Bevezetés a pedagógiai kutatás módszereibe* (pp. 171-211).

Navracsics, J. (2007a). *A kétnyelvű mentális lexikon*. Balassi Kiadó, Budapest.

Navracsics, J. (2007b). Word classes and the bilingual mental lexicon. In Navracsics, J., Lengyel, Zs. (Eds.). *Applied linguistic and psycholinguistic perspectives* (pp. 17-39) Clevedon, Multilingual Matters,

Navracsics, J. (2008). A kétnyelvűség pszicholingvisztikájáról - dióhéjban. In: Á. Vámos and J. Kovács, ed., *A két tanítási nyelvű oktatás elmélete és gyakorlata 2008-ban*, (pp.43-60). Eötvös József Könyvkiadó, Budapest.

Navracsics, J. (2010). *Egyéni kétnyelvűség*. Szegedi Tudományegyetem, Juhász Gyula Pedagógiai Kiadó. Szeged.

Navracsics, J., & Molnár, C. (2017). Multilingualism, Teaching, and Learning Foreign Languages in Present-Day Hungary. | *IRJE |Indonesian Research Journal in Education*|, *1*(1), 29-42.

Nikolov, M. (2011). Az angol nyelvtudás fejlesztésének és értékelésének keretei az általános iskola első hat évfolyamán. *Modern Nyelvoktatás*, *17(*1), 9-31.

Otero, T., & Barker, L. (2013). The Frontal Lobes and Executive Functioning. *Handbook Of Executive Functioning*, 29-44.

Paap, K., Anders-Jefferson, R., Mason, L., Alvarado, K. & Zimiga, B. (2018). Bilingual Advantages in Inhibition or Selective Attention: More Challenges. *Frontiers In Psychology*, *9*.

Paap, K., Johnson, H. & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265-278.

Patra, A., Bose, A., & Marinis, T. (2020). Lexical and Cognitive Underpinnings of Verbal Fluency: Evidence from Bengali-English Bilingual Aphasia. *Behavioral Sciences*, *10*(10), 155.

Pelham, S., & Abrams, L. (2014). Cognitive advantages and disadvantages in early and late bilinguals. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *40*(2), 313-325.

Pérez Cañado, M. (2018). CLIL and Educational Level: A Longitudinal Study on the Impact of CLIL on Language Outcomes. *Porta Linguarum Revista Interuniversitaria De Didáctica De Las Lenguas Extranjeras* 29, 51-70.

Pérez-Cañado, M. (2012). CLIL research in Europe: past, present, and future. *International Journal of Bilingual Education And Bilingualism*, *15*(3), 315-341.

Pfenninger, S. E. & Singleton, D. (2019). Making the most of an early L2 starting age. *Journal of Language Teaching for Young Learners 1*(2), 111-138.

Pino Escobar, G., Kalashnikova, M. & Escudero, P. (2018). Vocabulary matters! The relationship between verbal fluency and measures of inhibitory control in monolingual and bilingual children. *Journal Of Experimental Child Psychology*, *170*, 177-189.

Prievara, T. (2015): *A 21. századi tanár*. Neteducatio Kft. Budapest.

Racsmány, M. (2004). *A munkamemória szerepe a megismerésben*. Akadémiai Kiadó, Budapest.

Racsmány, M. (2007): Az „elsődleges emlékezet" – a rövid távú emlékezés és a munka - memória elméletei. In CSÉPE, V., GYŐRI, M. & RAGÓ, A. (szerk.): *Általános pszichológia 2*. Osiris Kiadó, Budapest. 177–209.

Ratiu, I., Hout, M., Walenchok, S., Azuma, T., & Goldinger, S. (2017). Comparing visual search and eye movements in bilinguals and monolinguals. *Attention, Perception, & Psychophysics*, *79*(6), 1695-1725.

Robinson, G., Shallice, T., Bozzali, M. & Cipolotti, L. (2012). The differing roles of the frontal cortex in fluency tests. *Brain*, *135*(7), 2202-2214.

Sajtos, L., Mitev, A. & Pusztai, T. (2007). *SPSS kutatási és adatelemzési kézikönyv*. Alinea, Budapest.

Sántha, K. (2015). *Trianguláció a pedagógiai kutatásban*. Budapest: Eötvös József Könyvkiadó, Budapest.

Saville-Troike, M. & Barto, K. (2012). *Introducing second language acquisition*. Cambridge University Press.

Selwyn, N. (2017). *Education and Technology – Key Issues and Debates.* New York: Bloomsbury Academic.

Shao, Z., Janse, E., Visser, K. & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, 5.

Shishkin, E. & Ecke, P. (2018). Language Dominance, Verbal Fluency, and Language Control in two Groups of Russian–English Bilinguals. *Languages*, *3*(3), 27.

Singleton, D. & Pfenninger, S. (2018). L2 acquisition in childhood, adulthood and old age: Misreported and under-researched dimensions of the age factor. *Journal of second language studies 1*(2), 254-275

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. Psychological Research, 63, 289–298.

Sullivan, M., Poarch, G. & Bialystok, E. (2017). Why is lexical retrieval slower for bilinguals? Evidence from picture naming. *Bilingualism: Language And Cognition*, *21*(3), 479-488.

Tánczos, T., Janacsek, K. & Németh, D. (2014). A verbális fluencia-tesztek II. A szemantikus fluencia-teszt magyar nyelvű vizsgálata 5-89 éves korig. *Psychiatria Hungarica, 29* (2) 181-207.

Túri, Zs., Németh, D. & Hoffman, I. (2014). Nyelv és emlékezet. In *Pszicholingvisztika 1-2*. (pp. 743-776).

Trilling, B. & Fadel, C. (2013): *21st century skills*. Jossey-Bass, San Francisco.

Troyer, A. (2000). Normative Data for Clustering and Switching on Verbal Fluency Tasks. *Journal Of Clinical And Experimental Neuropsychology*, *22*(3), 370-378.

Troyer, A. K., Moscovith, M. & Winocur, G. (1997): Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138-146.

Troyer, A., Moscovitch, M., Winocur, G., Alexander, M., & Stuss, D. (1998). Clustering and switching on verbal fluency: the effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia*, *36*(6), 499-504.

Troyer, A., Moscovitch, M., Winocur, G., Leach, L., & Freedman, a. (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of The International Neuropsychological Society*, *4*(2), 137-143.

Váczi, D. (2018): Utak a változásba. Az önszabályozás fontossága a 21. században. In: Bognár, A., Havassy, A., Nemes-Nagy, K. E., Váczi, D. (2018). *A modern pedagógus nagykönyve.* (pp.13-44). Neteducatio, Budapest.

Van Mensel, L., Hiligsmann, P., Mettewie, L., & Galand, B. (2019). CLIL, an elitist language learning approach? A background analysis of English and Dutch CLIL pupils in French-speaking Belgium. *Language, Culture And Curriculum, 33*(1), 1-14.

Verspoor M., De Bot, K. & Xu X. (2015) The effects of English bilingual education in the Netherlands. *Journal of Immersion and Content-Based Language Education 3*(1) 4–27.

Verspoor, M., & Hong, N. (2013). A dynamic usage-based approach to Communicative Language Teaching. *European Journal Of Applied Linguistics*, *1*(1).

Verspoor, M., Schmid, M. S. & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21(3), 239-263.

Weinreich, U. (1953): *The languages in contact*. The Hague, Mouton.

Wulff, S., & Ellis, N. (2018). Usage-Based Approaches to Second Language Acquisition. In D. Miller, F. Bayram, J. Rothman & L. Serratrice, *Bilingual Cognition and Language: The state of the science across its subfields* (pp. 37-56). John Benjamins Publishing Company.

Yu, Z. & Schwieter, J. W. (2018) Recognizing the Effects of Language Mode on the Cognitive Advantages of Bilingualism. *Frontiers in Psychology, 9*.

Zeng, Z., Kalashnikova, M., & Antoniou, M. (2019). Integrating bilingualism, verbal fluency, and executive functioning across the lifespan. *Journal of Cognition and Development, 20*(5), 656–679.

Further resources

Attard Montalto, S., Walter, L., Theodorou, M., & Chrysanthou, K. (2016). *The CLIL Guidebook*. Languages. dk. Retrieved 11 April 2021, from https://www.languages.dk/archive/clil4u/book/CLIL%20Book%20En.pdf

Delgado-Álvarez, A., Matias-Guiu, J., Delgado-Alonso, C., Hernández-Lorenzo, L., Cortés-Martínez, A., & Vidorreta, L. et al. (2021). Cognitive Processes Underlying Verbal Fluency in Multiple Sclerosis. *Frontiers In Neurology*, *11*. Retrieved on March 11, 2021 from https://doi.org/10.3389/fneur.2020.629183

Friesen, D., Latman, V., Calvo, A., & Bialystok, E. (2015). *Attention during visual search: The benefit of bilingualism*. Retrieved 14 March 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4667808/.

Kroll, J. F., & Bialystok, E. (2013). Understanding the Consequences of Bilingualism for Language Processing and Cognition. *Journal of cognitive psychology (Hove, England)*, *25*(5), 10.1080/20445911.2013.799170. Retrieved on January 20, 2020 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3820916/

Marope, M., Griffin, P., & Gallagher, C. (2017). Future competences and the future of the curriculum. Geneva: International Bureau of Education/UNESCO. Retrieved on April 10, 2021 http://www.ibe.unesco.org/sites/default/files/resources/future_competences_and_the_future_of_ curriculum.pdf

Marsh, J. E., Hansson, P., Sörman, D. E. & Ljungberg, J. K. (2019). Executive Processes Underpin the Bilingual Advantage on Phonemic Fluency: Evidence From Analyses of Switching and Clustering. *Frontiers in Psychology*. Retrieved 21 May 2021, from https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01355/full

McGuinness, C. (2018). *Research-Informed Analysis of 21st Century Competencies in a Redeveloped Primary Curriculum*. Ncca.ie. Retrieved 10 April 2021, from https://ncca.ie/media/4434/research-informed-analysis-of-21st-century-competencies-in-a-redevloped-curriculum.pdf.

Mészáros, A. (2017). Végrehajtó funkciók specifikus nyelvi zavarban és fejlődési dislexiában. (2017). Retrieved on February 9, 2020, from https://docplayer.hu/154421314-Meszaros-andrea-vegrehajto-funkciok-specifikus-nyelvi-za-varban-es-fejlodesi-diszlexiaban.html

Mihály, I. (2009). Európai áttekintés a két tannyelvű oktatás tapasztalatairól. Retrieved on July 19, 2020 from www.ofi.oh.gov.hu

Mohai K. , Kálózi-Szabó Cs. & Rózsa S. (2016) : A végrehajtó funkciók adaptív mérésének lehetőségei 40-85 *Psychologia Hungarica Caroliensis. évf. 1. sz.* 2016. Retrieved on March 24, 2020. from https://epa.oszk.hu/02400/02497/00007/pdf/EPA02497_PsychologiaHungaricaCaroliensis_2016_01_040-085.pdf

Oravecz, Cs., Váradi, T. & Sass, B. (2014): The Hungarian Gigaword Corpus. In: *Proceedings of LREC 2014*, 2014. http://mnsz.nytud.hu

Trentinné, B. É. (2014) A kétnyelvű fejlesztés és a pedagógusképzés. Neveléstudomány. OktatásKutatás-Innováció, 3. pp. 89-108. Retrieved on May 7, 2021 from: http://nevelestudomany.elte.hu/downloads/2014/nevelestudomany_2014_3_89-108.pdf

Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PloS one*, *8*(10), e77770. https://doi.org/10.1371/journal.pone.0077770

Zelazo, P. D., Blair, C. B. & Willoughby, M. T.. (2016). Executive Function: Implications for Education (NCER 2017-2000). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education Retrieved November 19, 2019, from https://ies.ed.gov/ncer/pubs/20172000/pdf/20172000.pdf

List of referred websites

https://www.oktatas.hu/kozneveles/kerettantervek/2020_nat/iranyelvek_alapprogramok

https://www.oktatas.hu/kozneveles/meresek/idegen_nyelvi_meres/jogszabalyok

https://www.oktatas.hu/pub_bin/dload/kozoktatas/meresek/celnyelvi_meresek/Celnyelvi_meres_tartalmi_keret.pdf

https://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign_language_skills_statistics#cite_note-1

Appendix A

SZÜLŐI NYILATKOZAT

kutatásban való részvételről

Tisztelt Szülő, Gondviselő!

Sántha-Malomsoki Ágnes vagyok, a Többnyelvűségi Doktori Iskola hallgatója. A doktori munkám alapját képező kutatás témája a második nyelv tanulásával, eltérő mértékű használatával járó kognitív változások (figyelem, nyelvi szerveződés) és a feltételezett összefüggések feltárása. A kutatásban tipikusan fejlődő általános iskolai hetedik és nyolcadik évfolyamos tanulók vesznek részt. A várt kutatási eredmények hozzájárulhatnak a korai és kései kétnyelvűség működési elvének megismeréséhez és az eredmények tudományos értelmezéséhez.

Gyermeke kutatásban való részvételéhez kérem az Ön beleegyezését. A részvétel önkéntes és névtelen. A tanulók adatait és eredményeit bizalmasan kezelem, azok Ön számára hozzáférhetőek, harmadik személy számára azonban nem elérhetőek. A tesztek felvétele három tanórát vesz igénybe, melynek lebonyolítása az intézménnyel előre egyeztetett időpontban történik.

Kérem, a megfelelő rész aláhúzásával szíveskedjen jelölni, amennyiben hozzájárul, illetve nem járul hozzá gyermeke kutatásban való részvételéhez.

Kijelentem, hogy gyermekem (név: ________________________________, a ________________________________ iskola tanulója a kutatásban

részt vehet nem vehet részt

____

Szülő/gondviselő aláírása

Együttműködését és beleegyezését előre is köszönöm.

Sántha-Malomsoki Ágnes

angoltanár

Székesfehérvár, 2019. szeptember

Appendix B

NYELVISMERETI KÉRDŐÍV (LEAP-Q)

Az alábbi kérdőív az általad használt nyelvekkel kapcsolatos kérdéseket tartalmaz. A kérdésekre nincsenek jó vagy rossz válaszok. A válaszaid egyediek lesznek, mert a te saját szokásaidat tükrözik. Kérlek, olvasd el a kérdéseket és a rád leginkább megfelelő válaszokat add vagy jelöld.

KÓD:		Születési idő		Dátum:	
Életkor:				Fiú ☐	Lány ☐

(1) Kérlek, sorold fel az összes nyelvet, amelyet használsz. Az általad <u>legjobban ismert</u> nyelv legyen az első.

1		2		3	

(2) Kérlek, sorold fel az összes nyelvet, amelyet használsz, az elsajátítás <u>időbeni sorrendje</u> alapján. (Az anyanyelved legyen az első.)

1		2		3	

(3) Kérlek, becsüld meg (%-ban), hogy az általad felsorolt nyelvek milyen mértékben vannak jelen a mindennapjaidban. (A százalékok 100%-ot adjanak ki.)

Nyelvek	1		2		3	
%						

(4) Kérlek, írd le, hogy az általad megadott nyelvek országában töltöttél-e hosszabb időt, s ha igen, mennyit:

	év	hónap	az életkorod ekkor
Nyelv 1			
Nyelv 2			
Nyelv 3			

(5) **Ha egy téged érdeklő szöveg, a számodra ismert valamennyi nyelven elérhető, melyik nyelven <u>olvasnád/hallgatnád</u>? (A százalékok 100%-ot adjanak ki.)**

Nyelvek	1		2		3	
%						

(6) Ha egy olyan személlyel szeretnél <u>beszélgetni</u>, aki az általad ismert valamennyi nyelvet hozzád hasonlóan ismeri, milyen arányban választanál a nyelvek között (A százalékok 100%-ot adjanak ki.)

Nyelvek	1		2		3	
%						

(7) Becsüld meg, hogy milyen szintet értél el, az alábbiakban:

NYELV 1	**semmilyen**	**nagyon gyenge**	**gyenge**	**elégséges**	**megfelelőnél kicsit**	**megfelelő**	**megfelelőnél kicsit job**	**jó**	**nagyon jó**	**kitűnő**	**tökéletes**
Beszéd											
Beszédértés											
Olvasás											
Írás											

NYELV 2	**semmilyen**	**nagyon gyenge**	**gyenge**	**elégséges**	**megfelelőnél kicsit**	**megfelelő**	**megfelelőnél kicsit job**	**jó**	**nagyon jó**	**kitűnő**	**tökéletes**
Beszéd											
Beszédértés											
Olvasás											
Írás											

NYELV 3	semmilyen	nagyon gyenge	gyenge	elégséges	megfelelőnél kicsit	megfelelő	megfelelőnél kicsit job	jó	nagyon jó	kitűnő	tökéletes
Beszéd											
Beszédértés											
Olvasás											
Írás											

(8) Jelöld a skálán, hogy az alábbi tevékenységeket milyen gyakran végzed, az általad ismert/használt valamennyi nyelven:

NYELV 1	soha	nagyon ritkán	néha	az idő felében	gyakran	többnyire	rendszeresen	szinte állandóan	állandóan
Olvasok									
Zenét hallgatok									
Filmet nézek									
Internetezem									
Chat-elek (írott szövegben)									
Tanulok									
Online játék másokkal									
Szövegírás (blog, napló, stb.)									
Skype-olás									
Családdal beszélgetek									
Információt keresek									
TV-nézés									

NYELV 2	soha	nagyon ritkán	néha	az idő felében	gyakran	többnyire	rendszeresen	szinte állandóan	állandóan
Olvasok									
Zenét hallgatok									
Filmet nézek									
Internetezem									
Chat-elek (írott szövegben)									
Tanulok									
Online játék másokkal									
Szövegírás (blog, napló, stb.)									
Skype-olás									
Családdal beszélgetek									
Információt keresek									
TV-nézés									

NYELV 3	soha	nagyon ritkán	néha	az idő felében	gyakran	többnyire	rendszeresen	szinte állandóan	állandóan
Olvasok									
Zenét hallgatok									
Filmet nézek									
Internetezem									
Chat-elek (írott szövegben)									
Tanulok									
Online játék másokkal									
Szövegírás (blog, napló, stb.)									
Skype-olás									
Családdal beszélgetek									
Információt keresek									
TV-nézés									

(9) **Becsüld meg, milyen fokú <u>akcentussal</u> rendelkezel az általad megjelölt nyelveken:**

	nincs akcentusom	**szinte nincs**	**nagyon enyhe**	**enyhe**	**kis**	**mérsékelt**	**észrevehető**	**jól észrevehető**	**erős**	**nagyon erős**	**rettenetesen erős**
NYELV1											
NYELV2											
NYELV3											

(10) **Kérlek, becsüld meg, hogy mások milyen gyakran érzékelik az akcentusod (ha van):**

	soha	**nagyon ritkán**	**néha**	**az idő felében**	**gyakran**	**többnyire**	**rendszeresen**	**szinte állandóan**	**állandóan**
NYELV1									
NYELV2									
NYELV3									

Appendix C

Structured interview questions

Kedves Résztvevő!

Kérlek, az alábbi kérdések őszinte megválaszolásával segítsd a munkámat. Megnyugtatlak, hogy nincsenek jó, rossz vagy elvárt válaszok, csupán egyediek, amelyek a TE egyéni véleményed tükrözik. A kérdőív kitöltése önkéntes és névtelen.

1. Miért tanulod az angol nyelvet?
   (Why are you learning English?)

2. Szerinted mi a legfontosabb az angoltanulásban?
   (What is the most important in learning English?)

3. Milyen a jó nyelvtanár szerinted?
   (What features a good language teacher?)

4. Milyen a nyelvtanároddal/nyelvtanáraiddal a kapcsolatod?
   (What features your relationship with your language teacher(s)?

5. Mennyire találod élvezetesnek az angolórákat? Miért?
   (Are your English lessons enjoyable? Why? Why not?)

6. Milyen feladatokat kedvelsz az angolórán?
   (What tasks and activities do you prefer in the English lessons?)

7. Milyen feladatokat nem kedvelsz az angolórán?
   (What tasks and activities don't you like in the English lessons?)

8. Melyek az erősségeid az angoltanulásban?
   (What are your strenghts in learning English?)

9. Melyek a gyenge pontjaid az angoltanulásban?
   (What are your weaknesses in learning English?)

10. Miben fejlődtél idén sokat az angol nyelvet tekintve?
    (In what field of English language have you improved a lot?)

11. Ha angoltanár lennél, mit csinálnál másként az órán?
    (If you were an English teacher, what would you do differently in the lesson?)

Appendix D

Total number of different cluster types and mean cluster size

		CLIL group				Control group			
		Phonological clusters		Semantic clusters		Phonological clusters		Semantic clusters	
		Slight	Strict	Slight	Strict	Slight	Strict	Slight	Strict
ENGLISH	F	92	28	20	14	81	16	24	2
	A	56	15	10	10	26	5	5	2
	S	93	37	21	16	82	17	23	5
	FAS Σ	**241**	**80**	**51**	**40**	**189**	**38**	**46**	**9**
HUNGARIAN	K	55	31	57	19	88	34	49	14
	T	80	25	37	12	84	40	54	23
	A	36	13	73	13	42	17	83	13
	KTA Σ	**171**	**69**	**167**	**44**	**214**	**91**	**186**	**50**
TOTAL		**412**	**149**	**218**	**84**	**403**	**129**	**232**	**59**

Mean cluster size (phonological + semantic)						
	ENGLISH			**HUNGARIAN**		
	F	A	S	K	T	A
CLIL	2.21	1.63	2.22	2.24	2.42	2.22
Control	1.19	0.95	1.83	2.30	2.53	2.10

Appendix E

Total number of words in different word classes

All generated words (**ENGLISH**)	CLIL group (Σ=1681 words)			Control group (Σ=1326 words)		
	F	A	S	F	A	S
Abstract nouns	55	47	43	23	11	11
Concrete nouns	258	239	317	215	162	243
Verbs	129	48	129	130	22	132
Adjectives	115	43	97	75	19	70
Others	35	65	61	70	69	74
Total	592	442	647	513	283	530

All generated words (**HUNGARIAN**)	CLIL group (Σ=1917 words)			Control group (Σ=2201 words)		
	K	T	A	K	T	A
Abstract nouns	35	71	58	29	59	48
Concrete nouns	489	440	369	588	552	445
Verbs	85	90	27	79	108	65
Adjectives	55	43	35	54	43	34
Others	21	18	81	15	22	60
Total	685	662	570	765	784	652

Appendix F

Data of participants selected for the structured interviews

Socio-economic status

Participants' codes	Age	Glasses	Gender	Handed-ness	Mother's level of education	Father's level of education	Parents' marital status
CLIL High							
RK1SZE	14	no	female	right	tertiary	secondary	separated
RK1MD	14	no	female	right	secondary	secondary	married
RK1SZR	14	yes	female	left	tertiary	tertiary	married
Control High							
RSKE	14	no	female	right	tertiary	tertiary	married
ZTAK	14	no	female	left	tertiary	tertiary	married
NLPR	14	no	male	left	tertiary	secondary	separated
CLIL Low							
RK1GS	14	yes	male	right	secondary	tertiary	married
RK4BB22	13	yes	female	right	secondary	secondary	separated
RK2MB	14	no	female	right	secondary	secondary	married
Control Low							
IKRTGJ	13	yes	male	right	secondary	secondary	married
IKZA	13	yes	female	right	tertiary	secondary	separated
IKTT	13	no	female	right	tertiary	secondary	married

Appendix G

Data of participants selected for the structured interviews

(d2-R test scores)

	CLIL 'high'			Control 'high'		
	RK1SZE	RK1MD	RK1SZR	RSKE	ZTAK	NLPR
Concentration (standard point)	119	114	113	113	110	119
Accuracy (Errors %)	2.07	4.97	1.17	2.85	4.19	9.66

	CLIL 'low'			Control 'low'		
	RK1GS	RK4BB22	RK2MB	IKRTGJ	IKZA	IKTT
Concentration (standard point)	94	93	89	90	94	92
Accuracy (Errors %)	25.97	9.83	18.85	35.58	16.91	14.96

Appendix H

Phonemic fluency test resultsof participants selected for the structured interviews

Mean results (FAS: English test; KTA: Hungarian test)		CLIL 'High'		Control 'High'		Difference between the groups	
		FAS	KTA	FAS	KTA	FAS	KTA
ALL WORDS		37.3	31.3	23.3	41.6	1.6x (CLIL)	1.33x (Con)
CLUSTERS	Phonological Slight	4.6	2.3	3.6	3.6	1.27x(CLIL)	1.56x (Con)
	Phonological Strict	4	1.6	0.3	2.3	13x (CLIL)	1.43x (Con)
	Semantic Slight	2	4.3	1.6	3.6	1.25x (CLIL)	1.19x (CLIL)
	Semantic Strict	0.6	0.3	0.3	1	2x (CLIL)	3.3x (Con)
MEAN CLUSTER SIZE		2.6	2.46	2.18	2.38	1.19x (CLIL)	1.03x (CLIL)
SWITCHES	Cluster	6.3	3	3.3	6.3	1.09x (CLIL)	2.1x (Con)
	Hard	10.3	12.3	13.3	18.3	1.29x (Con)	1.48x (Con)
'ERRING'	Errors	0	0	1.3	0	-	-
	Repetitions	0	0.3	0.3	0.3	-	-
	Perseverations	0	0	0	0	-	-

Mean results (FAS: English test; KTA: Hungarian test)		CLIL 'Low'		Control 'Low'		Difference between the groups	
		FAS	KTA	FAS	KTA	FAS	KTA
ALL WORDS		24.3	28	20	24.6	1.22x (CLIL)	1.13x(CLIL)
CLUSTERS	Phonological Slight	2.3	3	3	2.6	1.30x (Con)	1.28x (CLIL)
	Phonological Strict	1.3	0.3	1	1.3	1.3x (CLIL)	4.3x (Con)
	Semantic Slight	0.3	1.3	1	0.6	3.3x (Con)	2.16x (CLIL)
	Semantic Strict	0.6	0	0.3	0	2x (CLIL)	-
MEAN CLUSTER SIZE		2.3	1.63	1.86	2.18	1.24x (CLIL)	1.38x (Con)
SWITCHES	Cluster	3	2	1.3	1	2.03x (CLIL)	2x (CLIL)
	Hard	14.6	20	11.3	15.3	1.29x (CLIL)	1.3x (CLIL)
'ERRING'	Errors	1.6	0.3	1.3	0.6	1.23x (CLIL)	2x (Con)
	Repetitions	0	0	0	0	-	-
	Perseverations	0	0	0	0	-	-

Appendix I

‘High’ learners’ test results by which they were selected for the structured interviews

		CLIL ‘HIGH’						Control ‘HIGH’					
		RK1SZE		RK1MD		RK1SZR		RSKE		ZTAK		NLPR	
		FAS[1]	KTA[2]	FAS	KTA	FAS	KTA	FAS	KTA	FAS	KTA	FAS	KTA
ALL WORDS		25	34	69	32	18	28	20	42	28	39	22	44
CLUSTERS	Phonological Slight	4	1	8	3	2	3	6	4	2	4	3	3
	Phonological Strict	1	3	10	1	1	1	0	2	1	4	0	1
	Semantic Slight	1	6	3	6	2	1	2	5	3	2	0	4
	Semantic Strict	1	0	1	0	0	1	0	1	0	0	1	2
MEAN CLUSTER SIZE		2.5	3.1	2.8	2.1	2.6	2.2	2.08	2.43	2.16	2.36	2.3	2.36
SWITCHES	Cluster	4	4	13	3	2	2	6	9	3	5	1	5
	Hard	10	9	15	12	6	16	7	15	17	17	16	23
’ERRING’	Errors	0	0	0	0	0	0	1	0	1	0	2	0
	Repetitions	0	1	0	0	0	0	1	0	0	1	0	0
	Perseverations	0	0	0	0	0	0	0	0	0	0	0	0

[1] English phonemic fluency test
[2] Hungarian phonemic fluency test

Appendix J

Low' learners' test results by which they were selected for the structured interviews

		CLIL 'LOW'						Control 'LOW'					
		RK1GS		RK4BB22		RK2MB		IKRTGJ		IKZA		IKTT	
		FAS[1]	KTA[2]	FAS	KTA	FAS	KTA	FAS	KTA	FAS	KTA	FAS	KTA
ALL WORDS		17	18	25	31	31	35	17	22	18	30	25	22
CLUSTERS	Phonological Slight	1	1	3	2	3	4	2	4	2	2	2	2
	Phonological Strict	1	0	1	0	2	1	0	1	0	0	3	3
	Semantic Slight	0	1	0	2	1	1	1	1	2	0	0	1
	Semantic Strict	1	0	1	0	0	0	0	0	0	0	1	0
MEAN CLUSTER SIZE		2.16	1.3	2.5	1.3	2.26	2.3	1.3	2.1	1.5	1.3	2.8	3.16
SWITCHES	Cluster	0	4	1	1	8	1	1	1	1	0	2	2
	Hard	10	13	14	23	20	24	10	11	14	25	10	10
'ERRING'	Errors	3	0	1	0	1	1	0	0	4	0	0	2
	Repetitions	0	0	0	0	0	0	0	0	0	0	0	0
	Perseverations	0	0	0	0	0	0	0	0	0	0	0	0

[1] English phonemic fluency test

[2] Hungarian phonemic fluency test

Appendix K

(LEAP-Questionnaire - Individual differences of selected groups in their L1)

LEAP-Q L1 (Hungarian)	Speaking skill	Listening skill	Reading skill	Writing skill	Reading	Listening to music	Watching films	Interne use	Chatting	Learning	Online gaming	Texting	Skype use	Family chat	Info search	Watching TV
CLIL ’high’																
RK1SZE	11	11	11	11	9	3	8	2	2	7	1	1	1	9	5	8
RK1SZR	11	11	11	10	9	7	8	2	8	9	1	1	1	8	6	8
RK1MD	8	8	8	8	4	2	2	6	8	7	8	9	9	9	6	2
CLIL ’low’																
RK1GS	9	10	9	10	3	1	5	5	7	5	3	2	3	8	3	2
RK4BB22	5	9	8	8	8	6	8	7	8	7	5	3	5	7	7	3
RK2MB	9	9	8	6	5	3	9	9	9	8	2	1	6	9	8	9
Control ’high’																
RSKE	8	8	9	8	9	1	6	7	4	5	1	1	1	9	9	9
ZTAK	10	10	10	10	8	7	6	4	2	3	1	3	1	3	3	3
NLPR	10	11	9	8	7	9	7	8	8	6	9	2	8	7	6	8
Control ’low’																
IKRTGJ	10	9	8	7	5	3	8	7	8	3	8	1	1	8	9	3
IKZA	9	11	10	9	6	8	6	8	8	7	3	4	1	7	8	8
IKTT	11	9	10	8	3	9	5	7	7	8	3	7	7	7	6	3

LEAP-Q L2 (English)	Speaking	Listening	Reading	Writing	Reading	Listening to music	Watching films	Interne use	Chatting	Learning	Online gaming	Texting	Skype use	Family chat	Info search	Watching TV
CLIL 'high'																
RK1SZE	8	8	8	5	8	9	7	6	2	5	1	1	1	7	8	8
RK1SZR	8	9	8	8	3	9	3	3	1	8	1	1	1	6	8	6
RK1MD	6	8	9	7	9	9	5	5	5	5	7	8	6	3	3	1
CLIL 'low'																
RK1GS	9	8	9	8	3	9	5	6	3	5	5	2	2	2	3	3
RK4BB22	5	9	8	8	3	8	3	7	4	7	3	2	5	2	6	1
RK2MB	6	8	8	6	5	6	3	4	3	5	1	1	2	2	5	2
Control 'high'																
RSKE	9	9	8	8	2	9	7	3	6	1	9	2	1	1	3	2
ZTAK	8	9	8	8	7	6	6	4	1	3	1	1	1	3	3	4
NLPR	8	8	8	8	6	8	8	7	3	6	6	3	2	2	3	3
Control 'low'																
IKRTGJ	3	4	7	5	4	8	3	5	3	5	2	1	1	2	5	3
IKZA	8	8	8	7	5	6	9	8	5	6	4	1	1	3	7	8
IKTT	8	8	8	8	7	8	3	7	3	8	3	3	3	2	3	2