

*The misunderstood variable:  
Age effects as a function of type of instruction*

Simone E. Pfenninger

University of Zurich, Switzerland

[simone.pfenninger@es.uzh.ch](mailto:simone.pfenninger@es.uzh.ch)

Abstract

This study was designed to investigate the effects of age of onset and type of instruction on ultimate EFL attainment at the end of the period of normal schooling in Switzerland, measured in terms of written fluency, complexity, morphosyntactic accuracy, vocabulary size, and listening skills. Data were gathered from four groups of 18-year-old Swiss German learners of English: 50 were early starters who had attended an immersion (CLIL) program in elementary school and who continued CLIL in secondary school (EARLY CLIL), 50 had followed the same elementary school program but then received traditional EFL instruction after elementary school (EARLY MIX), 50 were late starters who began learning English immersively in secondary school, (LATE CLIL), while the other 50 attended a traditional EFL program in secondary school (LATE NON-CLIL). Results show that age of onset alone does not seem to be the distinguishing variable since early introduction of English in elementary school did not result in a higher level of proficiency when exposure to the language was limited to a few hours of class per week. The performance of the EARLY MIX participants was equaled and in certain areas significantly surpassed by the other groups, despite the additional five years of English study they had had in elementary school. The best results were found when early CLIL instruction was followed up by the use of English as an additional language of instruction in secondary school (EARLY CLIL group), which confirms the link between young starting age, implicit learning and long and massive exposure.

*Keywords:* second language learning, age effects in L2, input effects on L2, immersion, CLIL

## 1. Introduction

Fifteen years ago, Harley (1998) lamented the fact that no explanation had been provided for why in school settings “the additional time associated with an early headstart has not been found to provide more substantial long-term proficiency benefits” (p. 27). Despite the abundance of critical period studies, maturational state studies and ultimate attainment studies that have been carried out in the meantime, there are still unexplored issues regarding the amount and type of input needed for earlier starters to surpass later starters and be able to retain their learning advantages in the long term (e.g., Muñoz & Singleton, 2011). It is particularly important to revisit the achievements of prepuberty versus postpuberty learners of English as a foreign language (EFL) in various types of foreign language (FL) programs, as numerous educational authorities in Europe have recently brought forward the starting age of language instruction in elementary schools as a result of the “younger-is-better” view and the steady growth of English as a lingua franca. In Switzerland, as in many other countries, this has led to small amounts of FL instruction stretched over a rather long period of time. Interestingly, type of FL instruction was not matched to these new conditions and the new way individual learners learn after five years of CLIL instruction<sup>1</sup> in elementary school. While it is true that many Swiss secondary schools nowadays offer programs in which English is the language of instruction (so-called immersion programs), only a small percentage of the total number of Swiss students are actually fortunate enough to attend these programs (2% in 2005 according to Bürgi, 2007, p. 33). Instead, there has been a tremendous over-reliance on, and blind trust in, the age factor and the amount of time spent learning an FL, at the expense of the conditions of learning.

In light of their rather unimpressive impact, early FL programs are currently under scrutiny and the question has arisen as to how an earlier age of onset of acquisition (AoA) can be exploited more effectively. This is a question of considerable theoretical and practical significance since it is at the heart of the debates revolving around age as one of the most powerful and misunderstood variables in the research on FL learning and teaching, and it is integral to designing effective FL pedagogy (see, e.g., Larson-Hall, 2008). A major goal of this study, then, is to analyze learning outcomes in relation to AoA and instructional treatment by comparing 200 senior high school students with different constellations of FL learning. The issue at stake here is not to identify the ultimate attainment of FL learners but to analyze how the length of mandatory instructional time can be optimized for long-term benefits to unfold before the end of secondary education, given that we know today that “the necessary

---

<sup>1</sup> For a definition of CLIL, see below.

length of the relevant period of instruction [to reach native-likeness] is not within the bounds of possibility" (Singleton & Skrzypek, 2014, p. 6). Such final state data are invaluable for ongoing studies in FL learning settings in that they afford unique perspectives on the limits of multiple language acquisition in a formal, instructional setting. The findings suggest that the devil seems to be in the continuity of type of instruction: If the education system does not provide continuity in the immersion program for students moving from primary to secondary levels, early starters cannot profit from the extended instructional period of five years.

## 2. Overview of the theoretical landscape

### 2.1. Age of onset of acquisition: A misunderstood variable

While there seems to be abundant research as well as anecdotal evidence to show that, typically, there is a relationship between age and success in L2 learning in a naturalistic setting (for a recent review, see DeKeyser & Larson-Hall, 2005), only few linguistic advantages have been found in beginning the study of an L2 earlier in a minimal input situation (e.g., Cenoz, 2009; García-Mayo & García-Lecumberri, 2003; Moyer, 2013; Muñoz, 2006; Singleton & Ryan, 2004). Indeed, recent findings have cast doubt on the overreliance on biological and strictly cognitive dimensions of SLA in a formal instructional setting. Even though there is a group of researchers who suggest that younger learners are at an advantage as regards ultimate levels of attainment (e.g., Harley, 1986; Patkowski, 1980), the results of most classroom studies in the literature suggest that while biological age is important, much of its effect is the consequence of its co-varying relationship with nonbiological factors (for a state-of-the-art review see Muñoz & Singleton, 2011). As a result, traditional approaches to the critical period hypothesis (CPH) have been challenged on several fronts, notably in terms of the following insights:

1. the lack of a unitary critical period and the absence of a clear-cut endpoint (e.g., Bialystok, 1997; Birdsong & Molis, 2001; DeKeyser, Alfi-Shabtay, & Ravid, 2010; Singleton, 2005);
2. the relatively early onset of decline in L2 learning (possibly as early as age 6), which is subject to high variation among learners (e.g., DeKeyser et al., 2010; Singleton & Ryan, 2004);
3. the learning rate advantage of late starters over very long periods of time, even when differences in cognitive development disappeared with age, for example, when early starters reached puberty and the older learners' advantage was no longer due to their superior cognitive development (e.g., Muñoz, 2008, 2011);

4. the numerous accounts of late beginners who were able to reach native-like proficiency (e.g., Dörnyei, 2005; Moyer, 1999; Muñoz, 2006, 2008; Muñoz & Singleton, 2007);
5. the important roles of contextual factors (e.g., experience in the L2, type of instruction, quality of L2 input; see, e.g., Flege, 2009; Moyer, 2004) and individual factors such as L1 literacy skills, aptitude, motivation, sense of belonging, individual learning styles and strategies, self-efficacy, willingness to communicate, personality traits, knowledge of previous languages, and so on, and their interaction with age effects in SLA (e.g., Cummins & Swain, 1986; Dörnyei, 2005, 2009; Dörnyei & Ushioda, 2009; Lightbown, 2003; Moyer, 2004, 2013).

This list illustrates nicely that the age notion is a micro-variable that cannot be isolated from other co-occurring factors and that the lack of success that has been reported for early-entry programs has to be attributed to a variety of factors. The focus of this study is to take the discussion mentioned in 5 above a step further by focusing on the role and impact of contextual factors in relation to AoA through a comparison of learners enrolled in formal learning settings who receive different amounts and types of input.

## 2.2. Making the most of instructional time

Thus far, the L2 learning context has rarely been included as an important factor in the discussion of the CPH (Muñoz, 2006, p. 6), even though the type of instruction learners receive plays a decisive role in formal instructional settings since it determines, for instance, the quality and quantity of the input the students encounter and the variety and amount of practice opportunities they receive. It appears that age effects may not only differ according to whether the learning context provides learners with unlimited exposure to the target language (as in naturalistic language learning settings) but also whether exposure to the language is limited to a great extent (as in FL learning settings) or to some extent (as in school immersion settings) (Llanes & Muñoz, 2013).

In immersion programs, a number of content subjects (e.g., maths) are taught through the FL (but see Cenoz, Genesee, & Gorter, 2014 for a discussion of the wide scope of experiences encompassed by immersion type provision). Due to the higher amount and intensity of exposure to the FL, on the one hand, and the opportunities for engaging in authentic and meaningful interaction in real-life contexts, on the other, immersion students have traditionally been found to be highly successful in comparison with students who have received regular FL instruction, that is, instruction that focuses primarily on language learning and is

restricted to separate, limited periods of time or so-called “minimal input” of no more than four hours of instruction per week (Larson-Hall, 2008, p. 36). There is a general agreement in the immersion literature that students in concentrated programs generally acquire higher levels of proficiency in the L2 than students in programs with normally spaced L2 units of contact: “One or two hours a week – even for seven or eight years – will not produce advanced second language speakers,” write Lightbown and Spada (1993, p. 113). A large part of this problem is also the risk that students will have difficulty in seeing any progress over time, which might lead to frustration (Pfenninger, 2012; Tragant, 2006). Thus, students who have intensive exposure to the FL appear to have an advantage over those whose instruction is thinly spread out over a longer period of time (Muñoz, 2012; Netten & Germain, 2004; Serrano & Muñoz, 2007). In Switzerland, the longitudinal studies by Bürgi (2007) and Elmiger, Näf, Oudot, and Steffen (2010) uniformly found that late immersion students clearly outperform students in regular FL programs on a range of FL measures, while at the same time achieving the same levels of competence in academic domains, such as mathematics or science, as comparable students in L1 programs. However, while immersion students have been found to attain high levels of receptive skills, fluency and complexity in their FL (e.g., Swain & Lapkin, 2001), linguistic accuracy and lexical precision are said to cause more difficulties (e.g., de Graaff & Housen, 2009; Netten & German, 2004).

With regard to the age factor, there are studies that show that early immersion students obtain better results than late immersion students (e.g., Wesche, Toews-Janzen, & MacFarlane, 1996). However, the “older-is-better” trend has also been found in full immersion programs (Genesee, 2004). For instance, Lapkin, Swain, Kamin, and Hanna’s (1980) results revealed that the only skill in which early immersion students outperformed late immersion students was listening comprehension, in contrast to reading and writing skills, and lexical and grammatical knowledge. According to Genesee (2004), “bilingual programs that provide appropriate and continuous instruction can be effective with younger or older students; in other words, advanced levels of functional L2 proficiency can be acquired by students who begin bilingual education in the primary grades and by those who begin in higher grades” (p. 27).

### 2.3. The crux of implicit learning

Immersion programs and content and language integrated learning (CLIL) programs are known to foster so-called implicit learning<sup>2</sup> (de Graaff & Housen, 2009).

---

<sup>2</sup> According to R. Ellis (2005), implicit knowledge “is procedural, is held unconsciously, and can be verbalized only if it is made explicit. It is accessed rapidly and easily and thus is available for

In other words, just like (very young) children acquire their mother tongue rather unconsciously, unaware that they are actually learning it, immersion programs are designed to imitate this process by focusing on meaning and content rather than form (see Hulstijn, 2002, p. 204ff.). Whilst most studies agree that immersion students develop good comprehension, confidence and fluency in the target language (see discussion above), research supports the argument that an exclusive focus on meaning in comprehensible input, or language use, is not optimal when it comes to developing students' linguistic competence or bringing them to high levels of accuracy in their L2 (Genesee, 1987, 2004; Lyster & Ranta, 1997). In recent years, there has, thus, been an increased focus on the role of systematic language instruction along with a more explicit focus on linguistic forms ("focus on form," see Long & Robinson, 1998; Sharwood Smith, 1993). For instance, in their meta-analysis of studies that have examined alternative types of L2 pedagogy, Norris and Ortega (2000) found that instruction with an explicit focus on form was more effective than instruction with an implicit focus on form. Turnbull, Hart, and Lapkin (2003) suggest that the reason why initially early immersion students' achievement lags behind the performance of comparable students in the regular English program is a lack of formal instruction in English: "Once formal instruction begins, however, students show rapid gains in performance" (p. 8).

There are other well-known issues with implicit learning in the classroom in connection with maturational effects. Numerous researchers posit that earlier starters cannot gain the positive effects of an early start if there is insufficient input for the kind of implicit learning that is done by children while they are not cognitively ready yet for explicit methods of instruction (Dörnyei, 2005). Implicit learning occurs slowly and requires substantial exposure to L2 material, vastly more than a traditional input-impoverished classroom setting supplies (DeKeyser, 2000; DeKeyser & Larson-Hall, 2005; Hulstijn, 2002). What is more, school-based FL learning is typically predominantly explicit in nature (Singleton, 2005, p. 279).

Finally, many researchers (de Graaff & Housen, 2009; DeKeyser & Larson-Hall, 2005; Hulstijn, 2002) suggest that whereas younger learners learn implicitly, adults learn explicitly, that is, they have lost their ability to learn implicitly. For post-puberty L2 learners, DeKeyser (2000) proposes that implicit and explicit knowledge interact, in contrast to child L2 learners. Thus, the formula for the success of adolescent learners seems to be a combination of focus on form and focus on meaning, or explicit and implicit FL learning, respectively.

---

use in rapid, fluent communication" (p. 214). By contrast, explicit knowledge "is conscious and declarative and can be verbalized. It is typically accessed through controlled processing when learners experience some kind of linguistic difficulty in the use of the second language" (p. 214).

### 3. The study

#### 3.1. Research questions

The main goal of this study is to examine what extent type of instruction (regular EFL vs. CLIL) and AoA (early start vs. late start) have an effect on the absolute abilities of secondary school students at the end of the period of normal schooling, measured in terms of written fluency, complexity, morphosyntactic accuracy, and vocabulary size. Participants are 50 early starters who attended an immersion (CLIL) program in elementary school and who continued CLIL in secondary school (EARLY CLIL), 50 who followed the same elementary school program but then received regular EFL instruction after elementary school (EARLY MIX), 50 who were late starters who began learning English immersively in secondary school (LATE CLIL), and 50 who attended a regular EFL program in secondary school (LATE NON-CLIL). The following four research questions will be addressed:

1. Are the differences between the two age groups (early vs. late starters) significant at the end of secondary education in Switzerland?
2. Which age group benefits the most from which type of instruction (EARLY CLIL vs. EARLY MIX and LATE CLIL vs. LATE NON-CLIL)?
3. Is the interaction between the two independent variables (AoA and CLIL) significant?
4. Are the differences between the four learning constellations significant (EARLY CLIL vs. EARLY MIX vs. LATE CLIL vs. LATE NON-CLIL)? If yes, which constellation is most beneficial for the learning outcome of different starting age groups?

The EARLY CLIL group is expected to be at a learning advantage relative to the other three groups, owing to their early starting age in combination with long and intensive instruction and a combination of implicit and explicit instruction in secondary school: Not only did they have EFL classes (language classes) in secondary school, but also three school subjects which were taught in English. Furthermore, in line with findings in studies that looked into the beneficial effects of different learning contexts and different types of instruction (e.g., Llanes & Muñoz, 2013), immersion instruction in secondary school is hypothesized to be more beneficial for EFL acquisition than formal, explicit instruction. I would thus expect the groups who had CLIL in secondary school (EARLY CLIL and LATE CLIL) to outperform the other two groups (EARLY MIX and LATE NON-CLIL), while AoA has a neutral effect on the results.

### 3.2. Programs and participants

In order to examine the interaction between AoA and type of instruction, test results from 200 learners of English in Switzerland were obtained. I collected data in five schools in 12 English classes in the state system, ranging in size from 9 to 22 members. All the participants had similar characteristics: They were between 17 and 20 years old (mean 18;9), they came from similar socioeconomic backgrounds and did not take any private classes of English outside school. This design is supposed to ensure that no learner group can profit from cognitive advantages due to age (see Muñoz, 2008, p. 587ff. for a discussion of the impact of learners' chronological age). The participants are quite comparable to those in previous studies that analyzed age-related differences between early starters and late starters (e.g., Larson-Hall, 2008), in the sense that they represent an elite group.<sup>3</sup> Although some of the participants had different teachers, they followed a similar methodology and curriculum. Four of the 10 teachers involved were native speakers of English, while the other six were near-native speakers.

Because there was essentially no variation in age for the participants, it was decided to divide them into four groups according to age of onset and learning constellation in primary and secondary school instead of using correlational analysis, as described above: 50 EARLY CLIL, 50 EARLY MIX, 50 LATE CLIL, and 50 LATE NON-CLIL. Note that the early starters (EARLY CLIL and EARLY MIX) and the late starters (LATE CLIL and LATE NON-CLIL) had dissimilar amounts of exposure; due to their earlier start, the EARLY CLIL and EARLY MIX had had access to greater instruction time. However, the early starters were not mixed in with late starters in the same class.

In the German-speaking cantons of Switzerland, all students are required to study Standard German (the primary language of literacy) as an L2, English as an L3 and French as an L4 throughout the period of their primary schooling.<sup>4</sup> The Swiss Conference of Cantonal Ministers of Education promotes an implicit learning approach at primary school level called content and language integrated learning (CLIL). In the CLIL classroom, the accent is placed on EFL sensitization, oral fluency, comprehension, cultural awareness, vocabulary and formulaic language, based on the hypothesis that younger children cannot attend to formal, explicit FL instruction to the same extent as older children because prepubertal learning is less reliant on analytic ability (e.g., N. Ellis, 2002). Strictly speaking, the CLIL program is not an immersion program. While activities are undertaken in English, these activities relate to the learning of the second language. As such this

---

<sup>3</sup> Students in Switzerland are screened in grade 6, at which time they are placed in different school tracks based on their abilities. In order to attend an academically oriented high school (university preparatory school), they have to pass an entrance test.

<sup>4</sup> Note, however, that in certain cantons, French is taught before English.



program is similar to the “intensive English programs” in Canada (see, e.g., Netten & German, 2004), albeit with considerably fewer hours of instruction a week (two 45-min lessons per week). However, the strong focus on meaning in comprehensible input and the communication of authentic messages resemble the main goal of immersion programs.

The partial immersion programs that the EARLY CLIL and LATE CLIL attended in secondary school consisted of three content subjects (mathematics, biology and history) taught through the FL (L3 English) in order to maximize the quantity of comprehensible input and purposeful use of English, in line with Swain’s (1985) output hypothesis and Long’s (1981) interaction hypothesis. Additionally, English is taught formally as a separate school subject. Thus, learners experience a combination of formal and informal learning, which offers them what seems to be an ideal opportunity to learn an FL in a classroom: a combination of explicit learning, or “focus on forms,” and implicit learning, or “focus on meaning,” to use Long and Robinson’s terms (1998). It is important to note that the Swiss education system does not automatically provide continuity in the immersion program for students moving from primary to secondary levels. Students voluntarily opt for immersion instruction. Since the demand for a place in an immersion class is usually larger than the actual number of places available, a student’s average school grade functions as a criterion in deciding who can join the program and who cannot. In this study, the immersion students did not have significantly better grades in English before they entered the program, but they might have been significantly more motivated to study English than the students who did not sign up for an immersion program (see also Bürgi, 2007, p. 79; Elmiger et al., 2010, p. 70). Finally, note that in this paper, the notion of CLIL will be used as a cover term for both CLIL and immersion, following Mehisto, Marsh and Frigols (2008, chapter 1).

Overall, the EARLY CLIL group spent an average of 1,770 hours learning English from grade 2 to grade 12, followed by the LATE CLIL with 1,330 hours, the EARLY CLIL with 1,170 hours, and the LATE NON-CLIL with 730 hours. Other recent studies of maturational effects in a classroom have used shorter periods (from 600 to 800 hours) in their longest-term comparisons (e.g., García Mayo & García Lecumberri, 2003; Larson-Hall, 2008; Muñoz, 2006).

### 3.3. Measures and analyses

In order to reliably determine the absolute abilities of the learners tested here, we need convergent evidence from multiple elicitation methods that have been proven to show age-related differences in previous research (e.g., Llanes & Muñoz, 2013; Muñoz, 2006). It is also important to include measures that have been shown to be generally related to IQ scores (e.g., literacy-related skills, see Genesee, 1976)

as well as others that have not (e.g., listening comprehension skills, see Ekstrand, 1977). To attain the above-mentioned aims, the following skills are assessed: productive and receptive vocabulary size, grammaticality judgments, listening comprehension, written fluency, syntactic complexity, and morphosyntactic accuracy. Measures of basic interpersonal communicative skills were not included since it was suggested that they were less sensitive to individual cognitive differences and to academic development (see, e.g., Muñoz, 2006, p. 8).

Participants were asked to write one L2 English composition on the pros and cons of (reality TV) talent shows, a topic that was deemed suitable for adolescents and had been found to elicit different semantic and syntactic contexts (Pfenninger, 2011, 2013a, 2013b). They were then subjected to a listening comprehension test, which had been standardized with a population of EFL learners in Switzerland. To measure receptive morphosyntactic knowledge, a written grammaticality judgment task (GJT) was administered, which has been found to be a reliable and valid instrument in critical period studies (e.g., DeKeyser, 2000; García Mayo, 2003; Larson-Hall, 2008; McDonald, 2006). The GJT used here was a version of McDonald's (2006) test of basic English morphosyntax, adapted and used by Pfenninger (2011, 2013a, 2013b), which included 49 items and 15 distracters designed to test judgments on word order in declarative sentences (4 items), adverb placement (8 items), negation (5 items), *yes/no*-questions (4 items), *wh*-questions (4 items), article usage (6 items), regular past tense (6 items), regular plural (6 items) and third-person singular marking (6 items). The reliability coefficient (KR-20) obtained was .90 for grammatical items and .95 for ungrammatical items. The main advantages of a GJT are the following: (a) Since free production tasks (e.g., essays) involve the risk of avoidance of uncertain uses of linguistic forms, it is important to consult more controlled data sources; (b) the GJT is a response task designed to measure the (subconscious) knowledge of the linguistic rules that constitute the learner's internal grammar (see, e.g., García Mayo, 2003, p. 97); thus, in order for the participants not to draw on their explicit FL knowledge, the task was timed in this study, with the students having a maximum of 11 minutes to make their judgments (approximately 10 seconds per sentence); (c) the GJT is more direct and economical than spontaneous speaking and writing (Larson-Hall, 2008, p. 42); (d) the judgments may reflect information about implicit knowledge (Bialystok, 1981); (e) the correction of errors reflects explicit, analyzed knowledge that represents consciously held insights about language (Bialystok, 1981); thus, the participants in this study were asked to correct any sentences they considered ungrammatical; and, finally (f) in naturalistic settings, late learners have been found to experience more difficulty in their grammaticality judgments than early learners, since "memory capacity, decoding ability and processing speed are deficient in late L2 learners" (McDonald, 2006, p. 383; see also Larson-Hall, 2008).

Vocabulary size was assessed through the Academic sections in Schmitt, Schmitt and Clapham's (2001) Versions A and B of Nation's Vocabulary Levels Test, which includes academic words from the Academic Word List (AWL; Coxhead, 2000), fitting in a broad range between the 2,000 level and the 10,000 level (Schmitt et al., 2001, p. 68). Since this test does not provide direct information about the ability to use the target words productively (Schmitt et al., 2001, p. 62), it was decided to add the Productive Vocabulary Size Test by Laufer and Nation (1999), which gives some indication of size of productive mastery, as readers are required to supply the appropriate missing words in a cloze test with short contexts. Muñoz (2006, p. 19) suggests that such tests are cognitively demanding since they require understanding of a text and readers have to draw on their pragmatic knowledge as well as grammatical, lexical and contextual knowledge. After the completion of the tasks, participants filled in a biodata questionnaire adapted from the Language Contact Profile (LCP; Freed, Dewey, Segalowitz, & Halter, 2004).

### 3.4. Method and procedure

Three testing sessions of 45 minutes each were conducted with each class during regular class time. The order in which the tests were administered varied so as to control for a possible lack of attention problems (see Muñoz, 2006, p. 16). Following Llanes and Muñoz (2013), written fluency was examined in terms of words per T-unit (W/TU), that is, one main clause and all of the dependent modifying clauses (see also R. Ellis & Barkhuizen, 2005), while syntactic complexity was examined using the clauses per T-unit (CL/TU) complexity ratio. Accuracy was examined by counting the morphosyntactic errors per T-unit (ERR/TU), notably omission (e.g., *he love singing*), overuse (e.g., *she cans again drives*), substitution (e.g., *many people singing*), ir/regularization (e.g., *he taked*), misformation (e.g., *he get's*), systematic and random misorderings<sup>5</sup> (e.g., *singer bad* or *John carefully rode his motorcycle*), and "other" errors (e.g., agreement errors) (see McDonald & Roussel, 2010; Pfenninger, 2011). Unmarked forms were regarded as either present (omission of third person singular -s) or past (omission of regular or irregular past tense) depending on the tense used for other verbs in the sentence. If the target word was the only verb in the sentence then the tense of the previous sentence was considered.

### 3.5. Results

In the following, I will present the comparative analyses I performed for inter-group differences, inspected with analysis of variance (ANOVA) and independent

---

<sup>5</sup> For a discussion of random versus systematic misorderings, see R. Ellis and Barkhuizen (2005).

samples *t* tests (a Bonferroni adjustment was applied and an alpha level of .05 was set), and I will discuss the comparison with respect to the influence of AoA and type of instruction. In order to simultaneously test the effects of AoA and type of instruction as well as the interaction between these two independent variables, multivariate analysis of variance (MANOVA) was applied.

Table 1 Means (and standard deviations) and ANOVA results for learning constellation

	EARLY CLIL	EARLY MIX	LATE CLIL	LATE NON-CLIL	<i>F</i>	<i>df</i>	<i>p</i>	$\eta^2$	Favorable
LC	18.56 (1.81)	12.7 (3.16)	15.86 (3.16)	12.4 (3.68)	46.39	3	< .001*	.43	EARLY MIX
PV	35.52 (7.06)	24.98 (7.16)	35.37 (7.64)	25.65 (7.52)	31.73	3	< .001*	.35	EARLY MIX, LATE CLIL
RV	56.42 (5.27)	49.85 (6.15)	55.44 (6.20)	48.84 (7.84)	17.84	3	< .001*	.23	EARLY MIX, LATE CLIL
W/TU	17.70 (3.43)	14.25 (2.32)	17.10 (4.47)	13.35 (2.57)	14.83	3	< .001*	.27	EARLY MIX, LATE CLIL
CL/TU	1.76 (0.37)	1.58 (0.22)	1.79 (0.30)	1.57 (0.25)	8.02	3	< .001*	.12	EARLY MIX, LATE CLIL
ERR/TU	0.51 (0.29)	0.53 (0.34)	0.50 (0.43)	0.54 (0.30)	0.10	3	.96	.00	n.s.
GJT	42.83 (2.07)	42.84 (2.84)	42.92 (2.41)	43.10 (2.5)	0.13	3	.94	.00	n.s.

Note. LC = listening comprehension; PV = productive vocabulary; RV = receptive vocabulary; W/TU = written fluency; words per T-unit; CL/TU = written syntactic complexity: clauses per T-unit; ERR/TU = written accuracy; morphosyntactic errors per T-unit; GJT = grammaticality judgment task.

\*Statistically significant at  $\alpha < .05$ .

Table 1 presents the mean scores, standard deviations and intergroup differences for the seven tests. It can be seen in Table 1 that, generally, there were significant intergroup differences in listening comprehension (LC), productive vocabulary (PV), receptive vocabulary (RV), fluency (W/TU) and syntactic complexity (CL/TU), with rather high effects (between  $\eta^2 = .12$  and  $.43$ ). To answer the first research question, which asks whether the differences between the two age groups (early vs. late starters) are significant, we have to have a closer look at the performances of the early starters (EARLY CLIL and EARLY MIX) versus the late starters (LATE CLIL and LATE NON-CLIL).

Table 2 *t*-test results for the two age groups (early vs. late)

Test	<i>t</i>	<i>p</i>	Favorable
LC	-2.75	.006*	early
PV	0.21	.837	n.s.
RV	-2.75	.789	n.s.
W/TU	-0.53	.600	n.s.

CL/TU	0.16	.870	n.s.
ERR/TU	-0.01	.992	n.s.
GJT	0.52	.605	n.s.

Note. \*Statistically significant at  $\alpha < .05$ .

As shown in Table 2, from among the seven measures, only the differences in scores in listening comprehension were significant. Thus, late starters caught up with early starters' achievements in all skill areas except for listening comprehension, for which an earlier AoA was more advantageous. As the side-by-side graphic comparison of the groups in Figure 1 shows, however, this is due to the fact that the early starters in the immersion program (EARLY CLIL) significantly outperformed all the other groups, including the late starters in the same program (LATE CLIL) ( $t = -5.25, p < .001$ ).

The stellar performance of the EARLY CLIL clearly influenced the overall score of the early starters (EARLY CLIL and EARLY MIX together) in this area. Their superior performance on listening skills corroborates previous evaluations of programs with rich L2 exposure which have found that an implicit approach improves students' receptive skills in their L2 (de Graaff & Housen, 2009, p. 735). There were no significant differences between the EARLY MIX and LATE NON-CLIL in listening skills or any other measure (see Table 7 in the Appendix). Furthermore, older starters did not show greater variation in their L2 performance, as Table 1 above shows (see also Pfenninger, 2011).

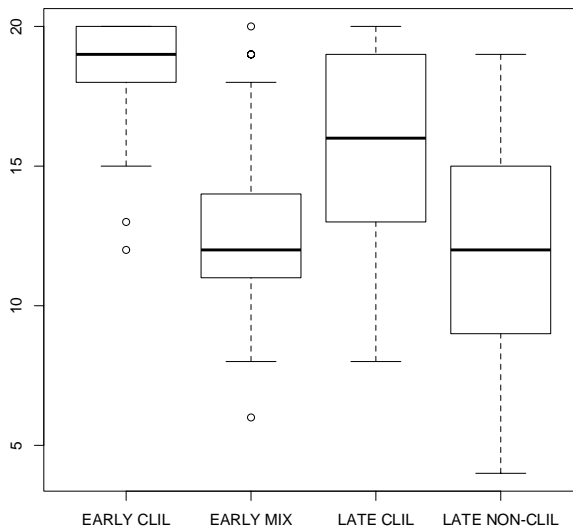


Figure 1 Listening scores (LC) by group (actual data points are overlaid on the boxplot and median lines are in bold)

My second research question asked which age group benefits the most from which type of instruction. It can be seen at first glance in Table 1 above that (a) immersion instruction was more beneficial than regular instruction alone, and (b) both early starters and late starters benefited from immersion. Evaluating immersion students against students who had studied EFL as a subject only, we find significant differences in receptive vocabulary knowledge ( $t = 4.99, p < .001$ ), productive vocabulary knowledge ( $t = 9.79, p < .001$ ), written complexity ( $t = 4.89, p < .001$ ) and fluency scores ( $t = 6.64, p < .001$ ) (see Table 8 in the Appendix). As expected and as shown in Tables 3 and 4, the EARLY CLIL and LATE CLIL participants had a significantly higher level of proficiency in English than the EARLY MIX and LATE NON-CLIL participants respectively, as reflected in higher values in most of the measures examined.

Table 3 *t*-test results for EARLY CLIL versus EARLY MIX

Test	<i>t</i>	<i>p</i>	Favorable
LC	11.49	< .001*	EARLY CLIL
PV	7.41	< .001*	EARLY CLIL
RV	5.73	< .0001*	EARLY CLIL
W/TU	10.30	< .0001*	EARLY CLIL
CL/TU	2.91	.005*	EARLY CLIL
ERR/TU	-0.28	.781	n.s.
GJT	-0.04	.968	n.s.

Note. \*Statistically significant at  $\alpha < .05$ .

Table 4 *t*-test results for LATE CLIL versus LATE NON-CLIL

Test	<i>t</i>	<i>p</i>	Favorable
LC	5.05	< .0001*	LATE CLIL
PV	6.41	< .0001*	LATE CLIL
RV	4.67	< .0001*	LATE CLIL
W/TU	3.77	< .0001*	LATE CLIL
CL/TU	4.02	.0001*	LATE CLIL
ERR/TU	-0.45	.652	n.s.
GJT	-0.37	.715	n.s.

Note. \*Statistically significant at  $\alpha < .05$ .

In the case of fluency, immersion students produced longer sentences, that is to say, sentences with a higher number of words per T-unit (W/TU). With respect to complexity (CL/TU), the LATE CLIL had the highest scores, followed by the EARLY CLIL, EARLY MIX, and LATE NON-CLIL, respectively. Immersion experiences are also beneficial for participants' lexical improvement (receptive and productive vocabulary), as can be seen in Figures 2 and 3.

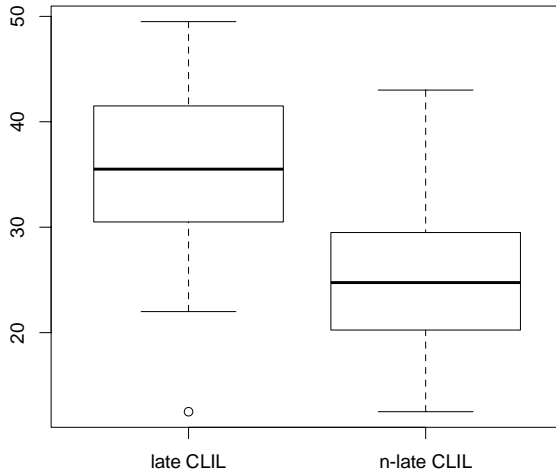


Figure 2 Productive vocabulary scores (PV)

While there were no significant differences between the performances of the EARLY MIX and LATE NON-CLIL (see Table 7 in the Appendix), the EARLY CLIL outperformed the LATE CLIL on one variable, namely listening comprehension ( $t = -5.25, p < .001$ ) (see Figure 1 above).

As can be seen in Table 1 and Figures 4 and 5, productive and receptive accuracy did not reach statistical significance. Although the LATE NON-CLIL perform slightly better on productive and receptive accuracy than the other groups, the differences are not statistically significant. To sum up, the ANOVA and  $t$ -test results for type of instruction indicate that immersion instruction was more beneficial than regular instruction for the improvement of written fluency, complexity, vocabulary size, and listening skills, but not as much for improving accuracy, as measured in this study.

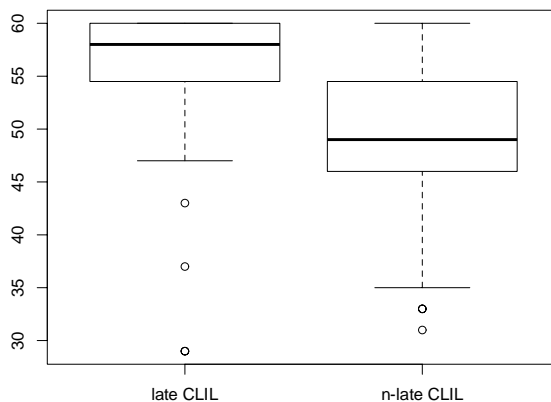


Figure 3 Receptive vocabulary scores (RV)

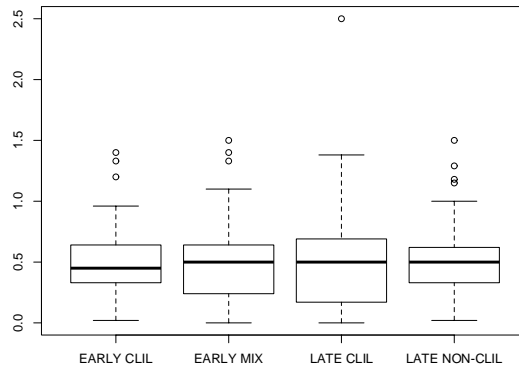


Figure 4 Productive accuracy scores (ERR/TU)

The results of the MANOVA tests concerning the impacts of AoA and type of instruction (the third research question) indicated that effects of instruction (i.e., late CLIL in secondary school) are significantly stronger than age effects for five of the seven dependent variables, as shown in Table 5. The follow-up analyses shown in this table show that effects of instruction were significant for listening comprehension, productive and receptive vocabulary, fluency and syntactic complexity, all with a high effect size. With respect to the impact of age, only listening comprehension reached statistical significance, with a small effect size, for which an earlier start was more advantageous. Concerning the interaction between AoA and type of instruction, MANOVA also indicated that the interaction was significant for listening comprehension, with a rather small effect, but not for any of the other dependent variables.

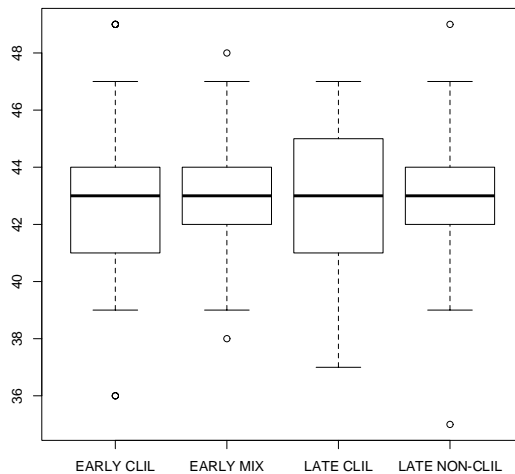


Figure 5 Receptive accuracy scores (GJT)



Table 5 Impacts of AoA and late CLIL and interaction between them (MANOVA)

Test	Effects of AoA			
	<i>F</i>	<i>df</i>	<i>p</i>	$\eta^2$
LC	11.91	1	.001*	.038
PV	0.06	1	.803	.000
RV	1.21	1	.273	.005
W/TU	0.33	1	.564	.001
CL/TU	0.03	1	.863	.000
ERR/TU	0.00	1	.992	.000
GJT	0.27	1	.607	.001
Test	Effects of late CLIL			
	<i>F</i>	<i>df</i>	<i>p</i>	$\eta^2$
LC	114.95	1	< .001*	.363
PV	94.98	1	< .001*	.326
RV	52.30	1	< .001*	.300
W/TU	43.58	1	< .001*	.182
CL/TU	23.69	1	< .001*	.108
ERR/TU	0.28	1	.600	.001
GJT	0.08	1	.775	.000
Test	AoA by late CLIL (interaction)			
	<i>F</i>	<i>df</i>	<i>p</i>	$\eta^2$
LC	7.89	1	.005*	.024
PV	0.16	1	.694	.001
RV	0.00	1	.982	.000
W/TU	0.56	1	.454	.002
CL/TU	0.34	1	.563	.002
ERR/TU	0.03	1	.872	.001
GJT	0.05	1	.819	.000

Note. \*Statistically significant at  $\alpha < .05$ .

To answer the last research question, the EARLY CLIL constellation seems to be more beneficial for learners in terms of the improvement of listening skills, but CLIL students generally presented the highest scores across most measures, irrespective of their starting age. This partly contradicts and partly supports the view that an early starting age produces long-term benefits when associated with greater time and massive exposure (e.g., Muñoz, 2008, p. 582).

### 3.6. Discussion

The results contribute to the growing body of research showing the existence of very proficient older starters. Both late-starting groups (LATE CLIL and LATE NON-CLIL) caught up with the EARLY CLIL group, who had benefited from longer exposure and therefore larger amounts of total input due to an earlier start. This supports the hypothesis that the initial advantage of older learners may last for several years in an input-impoverished environment and that it takes a

substantial accumulation of input to see advantages for an early start begin to show (Larson-Hall, 2008; Muñoz & Singleton, 2011; Singleton, 1995a, 1995b, 2005). The findings also support Muñoz and Singleton's (2011) view that age-related changes in language acquisition outcome might result from the impact of promotion factors other than a specifically language-focused critical period, such as deep and varied engagement with input in a classroom. Even though early learners (such as the EARLY CLIL and EARLY MIX in this study) may on average have greater potential than late starters due to their earlier AoA and the larger amount of cumulative input, this does not translate into better performance unless formal instruction in English in secondary school is supported by late immersion (see also Cenoz & Jessner, 2009, p. 132). Thus, where success is concerned, it is not so much about the age of onset or the length of the exposure to the L2, as has been suggested before. This leads us to the other independent variable investigated here: type of instruction.

Several factors might explain the EARLY CLIL participants' advantage: On the one hand, they received many more hours of (formal and informal) EFL instruction than any of the other groups. Furthermore, the immersion context fosters implicit learning, which is known to be more efficient than explicit mechanisms (DeKeyser, 2000). It could also be posited that the EARLY CLIL participants benefited more than the EARLY MIX participants from the early age of onset and the extended learning period because there was no abrupt transition from implicit to explicit learning for the EARLY CLIL group upon entrance into secondary school. It becomes apparent from the participants' responses in the motivation and biodata questionnaires that the EARLY MIX group did not react favorably to the new formal, explicit pedagogical approach they were faced with at the beginning of secondary school, which might have led to a lag in achievement (see also Pfenninger, 2011). By contrast, the EARLY CLIL participants were able to continue CLIL in secondary school in addition to formal EFL instruction.

Even more surprising than the EARLY CLIL students outperforming students in nonimmersion programs (EARLY MIX and LATE NON-CLIL) is the finding that the LATE CLIL group had made significant progress in a variety of skill areas, to the extent that they were able to catch up to the performance of the EARLY CLIL group. Thus, it seems to be access to late CLIL, regardless of early instruction, that makes the difference here. The oral-based, communicative pedagogical approach used in CLIL programs in secondary school could explain the significant differences in productive and receptive vocabulary knowledge and written complexity and fluency between the students who were immersively educated in secondary school (EARLY CLIL and LATE CLIL) and the traditionally instructed participants (EARLY MIX and LATE NON-CLIL), irrespective of their AoA. The success of the LATE CLIL is yet another indicator that instruction seems capable of overriding the

age factor in a classroom setting. However, in the absence of a control group who would have merely received more intense formal instruction, it cannot be ascertained beyond doubt that the combination of instruction and communicative exposure constitutes the optimal mix, since type of instruction and intensity of instruction became confounded (see discussion below).

The results also confirm the positive effects of form-focused instruction on acquisition (see studies cited in my review of the extant literature), that is, the effectiveness of explicit instruction on students' acquisition and use of specific morphosyntactic features of English. The tests assessing productive and receptive morphosyntactic accuracy revealed that students in immersion programs with more exposure to the target language do not always outperform students with less exposure, suggesting that simply extending exposure to and functional use of the target language do not necessarily lead to increased linguistic competence (Genesee, 1987, 2004). The lack of significant differences between all groups in relation to morphosyntactic accuracy might be due to the fact that the four groups practiced English grammar to the same extent. Since all the participants attended formal, explicit EFL instruction, they were required to read and write in English equally often and therefore paid great attention to accuracy.

#### 4. Conclusion

This paper sought to measure the magnitude of the effects of initial age of learning and type of instruction on a variety of EFL skills in an instructed setting. It has become clear in this study (and numerous earlier ones) that we cannot claim *per se* that the younger an L2 learner is when the L2 acquisition process begins, the more successful that process will be. The effect of additional instruction from an early age is only marginally seen in the learning constellation in which early CLIL instruction was followed up by the use of English as an additional language of instruction in secondary school (EARLY MIX group). On the one hand, this confirms previous observations that learners who experience intensive exposure to the FL in late immersion present similar levels of proficiency in the FL as children who have experienced more exposure to the FL in early immersion programs (see Genesee, 1987; Harley, 1986). On the other hand, it also shows that age matters in language learning, but only in the best-case scenario, that is, when it is "associated with *enough significant* exposure" (Muñoz, 2008, p. 591, my emphasis) and students receive a combination of explicit FL instruction and communicative exposure (see Graaff & Housen, 2009, p. 730). Thus, it was not my goal here to discuss whether there is a critical period or not, but rather to explore the factors, in addition to initial age of learning, that contribute to the high levels of proficiency achieved by some older starters.

Research on the interaction between age and type of instruction on the end point of acquisition of high school students has important implications for multilingual education when making decisions about (a) early instruction of different languages in elementary school and (b) later instruction through different languages in secondary school. Given the increasing number of early FL programs in Europe, state schools should consider offering more immersion programs or exchange programs at secondary level so that a larger number of students could actually profit from the earlier start of acquisition. The current Swiss system of formal education does not provide enough exposure to learners of English in order for the early starters to profit from the extended learning period. In partial immersion situations, input and use of the target language may also be limited but to a much lesser extent. Since immersion instruction appears to be effective with both elementary and secondary level students in most skill areas, irrespective of age of onset of acquisition, Genesee (2004) therefore suggests that in communities such as Germany (and I suggest also in communities that seek trilingual competence such as Switzerland), where monolingualism is the norm and other languages have no official status and/or are only used in restricted settings, introduction of immersion instruction in higher grades may be sufficient. However, because of the diversity of CLIL programs in Europe and the lack of conceptual clarity (see Cenoz et al., 2014, p. 257), it is difficult for researchers to provide a clear and detailed description of the CLIL classrooms/programs.

Some of the issues and questions that have emerged in the present study and that need to be addressed and considered in future research are:

1. One obvious limitation in this study is that since the CLIL groups not only had EFL classes (language classes), but also three school subjects which were taught in English, two variables were conflated at the same time in the CLIL groups: type of instruction and exposure (see Bruton, 2011; Cenoz et al., 2014). This is probably one of the most fundamental issues for CLIL researchers and will be difficult to resolve in the future.
2. One factor that can be—and has to be—controlled for in the future is aptitude, considering that “CLIL can attract a disproportionately large number of academically bright students” (Mehisto, 2007, p. 63).
3. It is also important as a next step to control for affective variables such as motivation, or, alternatively, to factor in motivation as yet another independent variable and to measure the effect of different motivational dispositions on the learning outcome and the interaction of motivation with AoA and type of instruction (see Lasagabaster, 2011; Pfenninger, 2014). Motivation is probably one of the most crucial factors in studies on the outcomes of CLIL, considering that in many European

countries CLIL programs are often not available to all students, which leads to a selection of students for these programs “who will be academically motivated to succeed in the FL (foreign language), as in other subjects” (Bruton, 2011, p. 524).

4. Finally, it also seems interesting to analyze which other input measures (length of instruction in years, number of curricular and extracurricular lessons, amount of time spent in a naturalistic immersion situation abroad, current informal contact with the target language) are strongly associated with long-term L3 performance (see Pfenninger & Singleton, 2014).

All these points call for further (critical) research that looks into the long-term learning benefits of early immersion instruction and the potential learning outcomes of children, adolescents, and adults in informal and formal L2 learning settings. Clearly, for educators, teachers and policy-makers, as well as for theorists, it is of compelling interest to know more about the end state of FL instruction. Despite the rich literature on maturational effects in formal instructional settings, which provides valuable clues to effective pedagogical practice, the interaction of age with other, contextual variables remains a controversial area of research.

## References

- Bialystok, E. (1981). The role of conscious strategies in second language proficiency. *The Modern Language Journal*, 65, 24-35. doi:10.1111/j.1540-4781.1981.tb00949.x
- Bialystok, E. (1997). The structure of age. In search of barriers to second language acquisition. *Second Language Research*, 13, 116-37. doi:10.1191/026765897677670241
- Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second language acquisition. *Journal of Memory and Language*, 44(2), 235-249. doi:10.1006/jmla.2000.2750
- Bruton, A. (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, 39, 523-532.
- Bürgi, H. (2007). *Im Sprachbad: Besseres Englisch durch Immersion. Eine Evaluation Zweisprachiger Ausbildungsgänge an drei Kantonalen Gymnasien in der Schweiz*. Bern: Hep.
- Cenoz, J. (2009). *Towards multilingual education: Basque educational research in international perspective*. Clevedon: Multilingual Matters.
- Cenoz, J., & Jessner, U. (2009). The study of multilingualism in educational contexts. In L. Aronin & B. Hufeisen (Eds.), *The exploration of multilingualism* (pp. 121-138). Clevedon: Multilingual Matters.
- Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, 35(3), 243-262. doi:10.1093/applin/amt011
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-38. doi:10.2307/3587951
- Cummins, J., & Swain, M. (1986). *Bilingualism in education*. London: Longman.
- De Graaff, R., & Housen, A. (2009). Investigating effects and effectiveness of L2 instruction. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 726-753). Oxford: Wiley-Blackwell.
- DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22(4), 499-533. Retrieved from <http://ruccs.rutgers.edu/~karin/550.READINGS/SYNTAX/CRITICAL.PERIOD/DeKeyser.pdf>
- DeKeyser, R. M., & Larson-Hall, J. (2005). What does the Critical Period really mean? In J. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88-108). New York: Oxford University Press.
- DeKeyser, R. M., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, 31, 413-438. doi:<http://dx.doi.org/10.1017/S0142716410000056>

- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.
- Dörnyei, Z., & Ushioda, E. (Eds.). (2009). *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters.
- Ekstrand, L. H. (1977). Social and individual frame factors in L2 learning comparative aspects. In T. Skutnabb-Kangas (Ed.), *Papers from the First Nordic Conference on Bilingualism*. Helsingfors: Universitetet.
- Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 141-188. doi:http://dx.doi.org/10.1017/S0272263102002024
- Ellis, R. (2005). Principles of instructed language learning. *System*, 33(2), 209-224. doi:10.1016/j.system.2004.12.006
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Elmiger, D., Näf, A., Reynaud Oudot, N., & Steffen, G. (2010). *Immersionsunterricht am Gymnasium: Eine Fallstudie zur Zwei-sprachigen Maturität in der Schweiz*. Bern: Hep.
- Flege, J. E. (2009). Give input a chance! In T. Piske T. & M. Young-Scholten (Eds.), *Input matters* (pp. 175-190). Clevedon: Multilingual Matters.
- Freed, B. F., Dewey, D. P., Segalowitz, N., & Halter, R. (2004). The language contact profile. *Studies in Second Language Acquisition*, 26, 349-356.
- García Mayo, M. (2003). Age, length of exposure and grammaticality judgments in the acquisition of English as a foreign language. In M. García Mayo, & M. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 94-114). Clevedon: Multilingual Matters.
- García-Mayo, M. P., & García-Lecumberri, M. L. (Eds.). (2003). *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.
- Genesee, F. (1976) The role of intelligence in second language learning. *Language Learning*, 26, 267-280. doi:10.1111/j.1467-1770.1976.tb00277.x
- Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Rowley, MA: Newbury House.
- Genesee, F. (2004). What do we know about bilingual education for majority language students? In T. K. Bhatia & W. Ritchie (Eds.), *Handbook of bilingualism and multiculturalism* (pp. 547-576). Malden, MA: Blackwell.
- Harley, B. (1986). *Age in second language acquisition*. Clevedon: Multilingual Matters.

- Harley, B. (1998). The outcomes of early and later language learning. In M. Med (Ed.), *Critical issues in early second language learning* (pp. 26-31). New York: Scott Foresman Addison Wesley.
- Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18(3), 193-223. doi: 10.1191/0267658302sr207oa
- Lapkin, S., Swain, M., Kamin, J., & Hanna, G. (1980). *Report on the 1979 evaluation of the Peel County late French immersion program, grades 8, 10, 11 and 12* (Unpublished report, University of Toronto, OISE).
- Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*, 24, 35-63. doi:10.1177/0267658307082981
- Lasagabaster, D. (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language Learning and Teaching*, 5(1), 3-18. doi: 10.1080/17501229.2010.519030
- Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, 16(1), 33-51. doi:10.1177/026553229901600103
- Lightbown, P. M. (2003). SLA research in the classroom/SLA research for the classroom. *Language Learning*, 28, 4-13. doi:http://dx.doi.org/10.1080/09571730385200151
- Lightbown, P. M., & Spada, N. (1993). *How languages are learned*. Oxford: Oxford University Press.
- Llanes, A., & Muñoz, C. (2013). Age effects in a study abroad context: Children and adults studying abroad and at home. *Language Learning*, 63(1), 63-90. doi 10.1111/j.1467 9922.2012.00731.x
- Long, M. H. (1981). Input, interaction and second language acquisition. *Annals of the New York Academy of Sciences*, 379, 259-278. doi:10.1111/j.1749-6632.1981.tb42014.x
- Long, M., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.
- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37-66. doi: 10.1017/S0272263197001034
- McDonald, J. L. (2006). Alternatives to the critical period hypothesis: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55, 381-401. doi:10.1016/j.jml.2006.06.006



- McDonald, J. L., & Roussel, C. C. (2010). Past tense grammaticality judgment and production in non-native and stressed native English speakers. *Bilingualism: Language and Cognition*, 13, 429-448. doi:10.1017/S1366728909990599
- Mehisto, P. (2007). What a school needs to consider before launching a CLIL program: The Estonian experience. In D. Marsh & D. Wolff (Eds.), *Diverse contexts-converging goals: CLIL in Europe* (pp. 61-77). Frankfurt: Peter Lang.
- Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL*. Oxford: Macmillan Education.
- Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition*, 21, 81-108.
- Moyer, A. (2004). *Age, accent and experience in second language acquisition. An integrated approach to critical period inquiry*. Clevedon: Multilingual Matters.
- Moyer, A. (2013). *Foreign accent. The phenomenon of nonnative speech*. Cambridge: Cambridge University Press.
- Muñoz, C. (2006). The effects of age on foreign language learning: The BAF project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1-40). Clevedon: Multilingual Matters.
- Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29, 578-596. doi:10.1093/applin/amm056
- Muñoz, C. (2011). Input in foreign language learning: More significant than starting age? *International Review of Applied Linguistics in Language Teaching*, 49(2), 113-133. doi: 10.1017/S0261444810000327
- Muñoz, C. (Ed.). (2012). *Intensive exposure experiences in second language learning*. Bristol: Multilingual Matters.
- Muñoz, C., & Singleton, D. (2007). Foreign accent in advanced learners: Two successful profiles. *EUROSLA Yearbook*, 7, 171-190.
- Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, 44(1), 1-35. doi:10.1017/S0261444810000327
- Netten, J., & Germain, C. (2004). Theoretical and research foundations of Intensive French. *Canadian Modern Language Review*, 60(3), 275-294. Retrieved from [http://www.mmecarr.ca/ICF/ICF\\_PDFs/Theory%20and%20Research.PDF](http://www.mmecarr.ca/ICF/ICF_PDFs/Theory%20and%20Research.PDF)
- Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417-528. doi:10.1111/0023-8333.00136
- Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, 30, 449-472. doi:10.1111/j.1467-1770.1980.tb00328.x

- Pfenninger, S. E. (2011). Age effects on the acquisition of nominal and verbal inflections in an instructed setting. *Studies in Second Language Learning and Teaching*, 1(3), 401-420. Retrieved from [https://repozytorium.amu.edu.pl/jspui/bitstream/10593/2310/1/SSLT%201\(3\)%20401420%20Pfenninger.pdf](https://repozytorium.amu.edu.pl/jspui/bitstream/10593/2310/1/SSLT%201(3)%20401420%20Pfenninger.pdf)
- Pfenninger, S. E. (2012). On the effectiveness of early implicit classroom learning – Evidence from morphology. *ELT Research*, 26, 30-32.
- Pfenninger, S. E. (2013a). On acquisition, age and articles in multilingual Switzerland. In J. Mihaljević Djigunović & M. Medved Krajnović (Eds.), *UZRT 2012: Empirical Studies in English Applied Linguistics* (pp. 22-35). Zagreb: FF Press.
- Pfenninger, S. E. (2013b). Quadrilingual advantages: Do-support in bilingual vs. multilingual learners. *International Journal of Multilingualism*, doi:10.1080/14790718.2013.782032.
- Pfenninger, S. E. (2014). All good things come in threes: Early EFL, motivation and CLIL in Switzerland. Manuscript in preparation.
- Pfenninger, S. E., & Singleton, D. (2014). *Beyond age effects: Facets, facts and factors of foreign language instruction in a multilingual state*. Bristol: Multilingual Matters. Manuscript in preparation.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88. doi: 10.1177/026553220101800103
- Serrano, R., & Muñoz, C. (2007). Same hours, different time distribution: Any difference in EFL? *System*, 35(3), 305-321. doi:10.1016/j.system.2007.02.001
- Sharwood-Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, 15, 165-179. doi:http://dx.doi.org/10.1017/S0272263100011943
- Singleton, D. (1995a). Introduction: A critical look at the Critical Period Hypothesis in second language acquisition research. In D. Singleton & Z. Lengyel (Eds.), *The Age Factor in second language acquisition: A critical look at the Critical Period Hypothesis* (pp. 1-29). Clevedon: Multilingual Matters.
- Singleton, D. (1995b). Second languages in the primary school: The age factor dimension. *Teanga: The Irish Yearbook of Applied Linguistics*, 15, 155-166. Retrieved from <http://files.eric.ed.gov/fulltext/ED407873.pdf>
- Singleton, D. (2005). The Critical Period Hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching*, 43(4), 269-285. doi: 10.1515/iral.2005.43.4.269
- Singleton, D., & Ryan, L. (2004). *Language acquisition. The age factor* (2nd ed.). Clevedon: Multilingual Matters.
- Singleton, D., & Skrzypek, A. (2014). Age and the classroom learning of additional languages. In M. Pawlak, J. Bielak, & A. Mystkowska-Wiertelak

- (Eds.), *Classroom-oriented research: Achievements and challenges* (pp. 3-13). Heidelberg: Springer.
- Swain, M. (1985). Bilingual education for the English-speaking Canadian. In J. Atlantis & J. Staczek (Eds.), *Perspectives on bilingualism and bilingual education* (pp. 385-398). Washington, DC: Georgetown University Press.
- Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 99-118). Harlow: Longman.
- Tragant, E. (2006). Language learning motivation and age. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 237-268). Clevedon: Multilingual Matters.
- Turnbull, M., Hart, D., & Lapkin, S. (2003). Grade 6 French immersion students' performance on large-scale reading, writing, and mathematics tests: Building explanations. *Alberta Journal of Educational Research*, 49(1), 6-23. Retrieved from <http://ajer.synergiesprairies.ca/ajer/index.php/ajer/article/view/354>.
- Wesche, M., Toews-Janzen, M., & MacFarlane, A. (1996). *Comparative outcomes and impacts of early, middle and late entry French immersion options: Review of recent research and annotated bibliography*. Toronto: OISE/UT Press.

APPENDIX

Table 6 *t*-test results for EARLY CLIL versus LATE CLIL

Test	<i>t</i>	<i>p</i>	Favorable
LC	-5.25	< .001*	EARLY CLIL
GJT	0.22	.824	n.s.
PV	-0.10	.919	n.s.
RV	-0.85	.397	n.s.
W/TU	-0.76	.449	n.s.
CL/TU	0.46	.645	n.s.
ERR/TU	-0.11	.909	n.s.

Note. \*Statistically significant at  $\alpha < .05$ .

Table 7 *t*-test results for EARLY MIX versus LATE NON-CLIL

Test	<i>t</i>	<i>p</i>	Favorable
LC	0.44	.661	n.s.
GJT	-0.49	.628	n.s.
PV	-0.46	.649	n.s.
RV	0.72	.471	n.s.
W/TU	-0.2	.842	n.s.
CL/TU	0.35	.727	n.s.
ERR/TU	-0.11	.910	n.s.

Table 8 *t*-test results for the two instruction types (LATE CLIL vs. N-LATE CLIL)

Test	<i>t</i>	<i>p</i>	Favorable
LC	-2.75	.006	CLIL
GJT	-0.29	.774	n.s.
PV	9.79	< .001*	CLIL
RV	4.99	< .001*	CLIL
W/TU	6.64	< .001*	CLIL
CL/TU	4.89	< .001*	CLIL
ERR/TU	-0.53	.598	n.s.

Note. \*Statistically significant at  $\alpha < .05$ .