

# **ALL GOOD THINGS COME IN THREES: EARLY ENGLISH LEARNING, CLIL AND MOTIVATION IN SWITZERLAND**

---

**Simone E. PFENNINGER**  
University of Zurich  
simone.pfenninger@es.uzh.ch

## **Abstract**

In this study, I examine the strength of the association between L3 English performance and starting age, on the one hand, and motivation and different types of provision of foreign language teaching, on the other, in Swiss learners of EFL with a long learning experience (between 6–11 years). Multilevel analyses were performed to investigate whether early starters in instructional settings achieve the same kind of long-term advantage as late starters and to examine how motivation and type of instruction (regular EFL instruction vs. Content and Language Integrated Learning or CLIL) factor into this process. Results show that starting age alone does not seem to be the distinguishing variable and that type of instruction and, above all, motivation are stronger predictors of L3 proficiency than starting age. Furthermore, qualitative analyses reveal a bi-directional causal link between CLIL and motivation and CLIL and learner outcomes. The study thus complements previous research by offering a critical empirical examination of age effects as well as CLIL outcomes and by investigating second-order interactions of individual difference variables and linguistic and contextual variables, which are still under-researched both in educational psychology and the study of second language acquisition.

*Key-words: age factor in SLA, foreign language learning, CLIL, immersion, motivation*

## **1. INTRODUCTION**

Education policy makers in many European countries tend to assume that age of instruction onset (AO) is the most important and robust predictor of success in foreign language learning in an instructional setting, “irrespective of what research findings suggest” (Mihaljević Djigunović, 2014: 420).<sup>1</sup> However, recent measures that have been implemented in Switzerland to improve students’ communicative skills and intercultural competence, such as the early teaching of English – learning English as a foreign language (EFL) starts as early as the age of eight or nine now in 15 out of 26 cantons in Switzerland (EDK 2014) – have

---

<sup>1</sup> To give an example for this line of argumentation: in 2003, the Bildungsrat of the Canton Zurich explained that early English was introduced because “younger learners are capable of acquiring and storing a language unconsciously, provided they are exposed to regular and rich input. Language skills that are stored in this manner will automatically be available to the learners later in life” (Bildungsratsbeschluss 18/3/2003, my translation). For a recent publication on language policy documents in Europe, the interested reader is referred to Nikolov and Mihaljevic Djigunovic (2011).

yielded rather disappointing first results, and it has become clear that the early teaching of English may not – cannot – be the sole course of action to improve students' English language competence. At this time it is particularly important to revisit linguistic and affective characteristics of early starters vs. late starters in various types of foreign language (FL) programs in Switzerland, as educational authorities in Europe have recently brought forward the starting age of language instruction in elementary schools, mainly as a result of the “younger-is-better” view and the steady growth of English as a *lingua franca*, although other reasons are also mentioned in official Swiss language policy documents, such as the political and cultural significance of the four national languages (German, French, Italian, Rhaeto-Romanic) on a national level, later learned languages (particularly French, see Haenni Hoti et al., 2011), the multilingualism requirement/goal in Europe, parental encouragement, globalization, integration and the world-wide network, and favorable attitudes to other languages, people, and cultures (see EDK 2014, 2015; Eurobarometer 2006; European Commission 1995). This has led to small amounts of second language (L2) instruction stretched over a rather long period of time, which may have an impact on students' motivation, especially in the long term (Lasagabaster, 2011: 13).

Around the same time as the Swiss Conference of Education Directors decided to lower the starting age of English instruction, they also started to implement Content and Language Integrated Learning programs, generally known as CLIL, in which three content subjects (such as mathematics or biology) are taught through the FL.<sup>2</sup> The introduction of CLIL in Switzerland reflected a general need in Europe to provide students with enhanced opportunities in school to acquire competence in additional languages (see Marsh, 2002). Since then, the very positive associations of CLIL (e.g. its perceived success and effectiveness) have attracted researchers, administrators, teacher educators, and teachers, particularly those in the field of English as an L2/FL (Cenoz, Genesee & Gorter, 2014: 247). Anton Näf, Emeritus Professor at the University of Neuchâtel, even went as far as to call CLIL programs in Switzerland “the egg of Columbus” (*Tages-Anzeiger* 27/1/2014) due to its potential to improve students' FL skills.

It is one of the main goals of this study to offer a critical empirical examination of age effects in interaction with CLIL in state educational institutions in order to better identify the strengths and weaknesses of different FL programs. However, any study of the outcomes of CLIL has to take into account one of the most crucial factors interacting with type of instruction, namely motivation, since in many European countries, CLIL programs are often not available to all students, which leads to a selection of students for these programs “who

---

<sup>2</sup> See below for a more detailed definition of CLIL.

will be academically motivated to succeed in the FL, as in other subjects” (Bruton, 2011: 524). Examining the impact of starting age, type of instruction and motivation in the same study may help us understand the relative importance of each of these factors for language use, an insight that has been impossible to gain in previous studies.

I would like to point out that due to the scope of this paper and the focus of this special issue, it was not possible to consider English in the broader multilingual context (e.g. in relation to French as an additional foreign language). The interested reader is referred to Pfenninger and Singleton (in prep) and Pfenninger and Singleton (submitted), where we analyze in detail the causes of and constraints on crosslinguistic influences in the Zurich system, including the French-English interaction and the socio-affective dimension (motivation, attitudes, awareness, anxiety, and learning strategies).

## **2. BACKGROUND TO THE STUDY**

### 2.1 AGE-BY-TREATMENT INTERACTION RESEARCH

Age-by-treatment interaction research has traditionally shown that different learning processes are at work at different ages, which explains the need for different “treatment”, both in the broad sense (exposure in a naturalistic setting vs. instruction in a classroom) and in the narrow sense (e.g. meaning-focused vs. form-focused instruction) (see DeKeyser, 2012). In this section, I will first focus on the macro level, that is, age \* treatment (context) interaction<sup>3</sup>, followed by a discussion of the micro level, i.e. age \* (instructional) treatment interaction.

Numerous classroom studies in Europe and indeed across the world (see, e.g., Al-Thubaiti, 2010 for Saudi Arabia; Muñoz, 2006, 2011 for the Basque Country; Larson-Hall, 2008 for Japan; Myles & Mitchell, 2012 for GB; Unsworth, de Bot, Persson & Prins, 2012 for the Netherlands, just to name a few) have found that there are no correlations between starting age and FL language outcomes in formal instructional settings, in contrast to the situation in naturalistic settings (for a recent review, see DeKeyser & Larson-Hall, 2005). The main goals of research in FL learning settings have been to examine FL outcomes as a function of the starting age, size and characteristics of older learners’ advantage for different language dimensions and after different amounts of exposure, and, more recently, the interplay of the age factor with social, affective and personal variables. Contextual factors, such as amounts and intensity of input (see, e.g., the collection in Muñoz, 2012a), high-quality input (e.g. Winitz, Gillespie & Starcev, 1995; Flege & Liu, 2001), range of contexts

---

<sup>3</sup> “\*” indicates interaction.

of L2 use (e.g. Moyer, 2004), and co-habitation with native speakers (e.g. Muñoz & Singleton, 2007; Kinsella & Singleton, 2014), have been shown to have a significant impact on learners' attainment (for a review see Muñoz & Singleton, 2011). In the following I will focus on amount and intensity of input, i.e. different types of provision of FL teaching in a classroom, notably CLIL vs. regular EFL instruction.

Launched in Europe in the 1990s by “a group of experts from different backgrounds” (Cenoz et al., 2014: 243), CLIL is “a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language” (Coyle, Hood & Marsh, 2010: 1; see Cenoz et al. 2013 for a description of the wide range of educational CLIL practices). The dual role of language and content thus means that proficiency is to be developed in both the non-language subject and the language in which it is taught (Lasagabaster, 2011) – although it is notoriously difficult to achieve a strict balance of language and content, which leads to “a lack of cohesion around CLIL pedagogies” (Coyle, 2008: 101; see also Cenoz et al., 2014; Mehisto, 2008; Pérez-Vidal & Juan-Garau, 2010). Since the definition of CLIL now also includes reference to partial immersion (Cenoz et al., 2014: 246; Maillat, 2010; see Pérez-Cañado, 2012, for an opposing view), the notion of CLIL will be used in the following as a cover term for both CLIL and immersion<sup>4</sup>.

A considerable amount of CLIL research has been carried out in intensive primary and secondary school classes in the last twenty years, and various benefits of CLIL have been pointed out, such as the following:

- (1) 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, particularly with respect to receptive skills (listening and reading), oral fluency, syntactic complexity, lexical range and confidence/risk-taking in the target language (e.g., Collins & White, 2011, 2012; Dalton-Puffer, 2007; Spada & Lightbown, 1989; Pfenninger, 2014; Ruiz de Zarobe & Jiménez Catalán, 2009; Serrano & Muñoz, 2007);
- (2) CLIL students have been reported to demonstrate better verbal and non-verbal communication skills, cognitive skills and divergent thinking than their non-CLIL counterparts (Vesterbacka, 1991);

---

<sup>4</sup> Note that ‘immersion’ is the term more commonly used in the Swiss context for what is done at grammar schools, while the label ‘CLIL’ dominates in state secondary schools.

- (3) The above benefits have emerged both when exposure has been concentrated and when it has been distributed across time in short intensive experiences (e.g., Collins & White, 2011, 2012);
- (4) CLIL is said to be able to minimize the role that individual differences, such as language learning aptitude, may play in more limited exposure situations (e.g., Collins & White, 2011, 2012);
- (5) CLIL increases exposure to the target language without taking up more time in an already crowded school timetable (e.g., Lasagabaster, 2011);
- (6) Content knowledge appears to remain on a par with that learned through the L1 (e.g., Admiraal, Westhoff & de Bot, 2006; see also Cummins, 1995; Genesee, 1987, 2004);
- (7) L1 skills are very similar both in CLIL classes and in non-CLIL classes (e.g. Seikkula-Leino, 2007; Vesterbacka, 1991);
- (8) Due to the higher exposure to the FL than in regular programs, CLIL programs are known to foster implicit learning,<sup>5</sup> which has been identified as a highly effective way of learning (Coyle, 2008; de Graaff & Housen, 2009; DeKeyser, 2000; Hulstijn, 2002);
- (9) Related to point (8), CLIL is age-appropriate in elementary schools, since younger children (e.g. in an early FL program) cannot attend to formal, explicit L2 instruction to the same extent as older children as prepubertal learning is less reliant on analytic ability (e.g., N. Ellis, 2002).

Of course there are numerous well-known issues with the implementation of CLIL in the classroom, particularly with implicit learning in connection with maturational effects, but I do not wish to go very deeply into this here (the interested reader is referred to Pfenninger, 2011, 2014; Pfenninger & Singleton, in prep.). The important point here is that AO-treatment interaction research shows more than the importance of starting age or a particular treatment. It can show why a treatment works best (or more precisely why sometimes it does and

---

<sup>5</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 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).

sometimes doesn't): due to the learning processes it involves, the treatment works well only with certain AO groups.

## 2.2. AGE \* (INSTRUCTIONAL) TREATMENT \* MOTIVATION INTERACTION

From the above discussion it has become clear that in order to have valid comparisons of the effect on learner outcomes in CLIL vs. non-CLIL classes, it is inevitable either to control for the motivational levels of the students or, preferably, to use motivation as yet another fixed effect in the statistical model. The theoretical framework for motivation in this study is based on the L2 Motivational Self System proposed by Dörnyei (2005, 2009), which hypothesizes that students' motivated learning behavior will be largely affected by three variables. The Ideal L2 Self, the person the learner wants to become, incorporates "traditional integrative and internalised instrumental motives". The Ought-to L2 Self, the side which wants to avoid punishment and meet expectations, incorporates "more extrinsic (i.e. less internalised) types of instrumental motives" (Dörnyei, 2009: 9).<sup>6</sup> A third component, L2 Learning Experience, covers the more immediate learning situation important to any study of L2 motivation in a classroom context (syllabus, teacher, etc.). Instrumentality is thus partly related to the Ideal L2 Self, particularly instrumentality with a promotion focus (Dörnyei, 2005: 30). Other forms of instrumentality (e.g. instrumentality with a prevention focus) may be more associated with the Ought-to L2 Self, the image of oneself which avoids punishment, i.e. external regulation.

As Dörnyei and Chan (2013: 439) point out, numerous studies in recent decades have confirmed the overall explanatory power of the L2 Motivational Self System, with the Ideal L2 Self in particular seen as a strong predictor of various criterion measures related to language learning, thus playing a substantive role in determining motivated behavior. For instance, research by Csizér and Lukács (2010) confirmed the seminal role of Dörnyei's conception of the Ideal L2 Self in predicting motivated learning behavior, and the paramount influence of the Ideal L2 Self on motivation. By contrast, it has often been suggested that the Ought-to L2 Self appears to have no significant impact on results (e.g. Csizér & Dörnyei, 2005: 29). However, Csizér and Lukács (2010: 12) argue that further research on, and reformulation of, the concept of the Ought-to L2 Self may clarify this aspect of the L2 Motivational Self System. What is more, age plays a role in the formation of selves: Kormos and Csizér (2008) found that secondary school students in Hungary scored lower values for the Ideal L2 Self than university students or adult workers, speculating that "students' self-

---

<sup>6</sup> Note that it is beyond the scope of this paper to discuss the problems surrounding conceptualizations of the self (see Mercer & Williams, 2014).

image is relatively stable, and because they have to acquire the L2 in adulthood, the L2 self is also under transformation at this stage” (2008: 346).

### 3. THIS STUDY

#### 3.1 RESEARCH QUESTION AND VARIABLES

On the basis of what has been discussed so far, the present study aims to enrich our knowledge of the effects of input in long-term FL learning by exploring which of the three predictors (starting age, type of instruction, motivation) has a stronger predictive power. The following summarizes the main research question:

- (1) What is the strength of the association between L3 performance and starting age, on the one hand, and type of instruction and motivation, on the other, in learners with a long learning experience (between 6–11 years)?

Individual differences factors in this study are AO and motivation. The context-level factor, CLIL, is hypothesized to influence individual EFL proficiency through its mediating effect on the association between individual differences factors and L2 proficiency. Although I hypothesize motivation and CLIL to have a positive effect and starting age to have a neutral effect on EFL proficiency at both the individual and contextual levels, how and to what extent individual- and contextual-level factors may interact with each other are open empirical questions.

#### 3.2 PARTICIPANTS: NESTING STRUCTURE

A total of 200 participants (89 males and 111 females) were clustered in 12 classes in five schools, mostly consisting of 10–20 learners, all of whom had similar characteristics: they were in grade 12 English classes in academically oriented secondary school, 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.<sup>7</sup> Through this clustering, participants were streamed into two different instruction types: on the one hand, there were students who were enrolled in CLIL programs (100 students in six classes) and, on

---

<sup>7</sup> The participants were drawn from a larger sample of intermediate/advanced learners of English (see Pfenninger & Singleton, in prep). Note that in this project an additional control group of 100 early starters in grade 7 (age 13) was recruited in 2014 in order to be able to obtain a realistic picture of the benefits of early English programs. These students belong to the fifth cohort of early English learners in the canton of Zurich.

the other, students who followed an EFL approach and who only had exposure to EFL in the traditional way (100 students in six classes). As mentioned above, students in CLIL classes received additional exposure to the foreign language: English classes as well as the school subject taught in English. Furthermore, they were divided into four groups according to age of onset and learning constellation in primary and secondary school: 50 of the participants were early starters who attended an immersion (CLIL) program in secondary school (EARLY CLIL), 50 had followed the same elementary school program but then received regular EFL instruction after elementary school (EARLY NON-CLIL), 50 were late starters who began learning English immersively in secondary school (LATE CLIL), while the other 50 attended a regular EFL program (LATE NON-CLIL). Note that the early starters (EARLY CLIL and EARLY NON-CLIL) 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 NON-CLIL had had access to greater instruction time. By the end of secondary school, the EARLY CLIL group spent an average of 1,770 hours learning English, followed by the LATE CLIL with 1,330 hours, the EARLY NON-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). However, the early starters were not mixed in with late starters in the same class. A biodata questionnaire was administered to collect biographical data and quantifiable information concerning their language learning experience (e.g., starting age, number of instructional hours in school, frequency of contact with L2 speakers, time spent abroad).

The school track under investigation here, which I refer to as ‘academically oriented secondary school’, represents the main – but not the only – university entry pathway. It is an elite and selective publicly funded school, representing one of three main secondary school tracks (the highest educational level). In the canton of Zurich, admission is based on students’ average grades and an entrance examination. The number of those taking the matura or maturité exam (i.e. the final graduation exam) has increased in recent years. Between 1986 and 2013 the percentage awarded this certificate almost doubled to 20 percent (<http://www.bfs.admin.ch/bfs/portal/de/index/themen/15/01/pan.html> on 12/7/2015). There are three main reasons why it was decided to assess the development of EFL skills of this group of learners:

- (1) This particular secondary school track is roughly equivalent to grammar schools, Baccalaureate schools and high schools in other countries in terms of length of



instruction (six years until graduation), institutional design (e.g. number and kinds of compulsory subjects, assessment of students, final certificate) and purpose (e.g. they do not lead to professional qualifications, but prepare students for tertiary level education programs). This is important for comparisons with related previous work in Europe and elsewhere.

- (2) Lower secondary levels, which only take three years, are not ideal to test for *long-term* effects of an early foreign language program. In age-related research, it is one of the most basic and most important tasks to identify predictors of short-term AND long-term FL attainment. Furthermore, it has been previously suggested that it takes a substantial accumulation of input to yield manifestations of advantages of an early start (e.g., Larson-Hall, 2008; Muñoz & Singleton, 2011; Singleton, 1995a, 1995b, 2005).
- (3) Assessing “good and motivated” learners<sup>8</sup>, who (ideally!) involve themselves in the language-learning process and take into account the demands that FL learning imposes, is not considered a limitation in this kind of study: strong learners can provide key data on the effectiveness of a new FL program and yield revealing results in search of influential factors in the process of FL learning (see, e.g., Muñoz, 2014). The insights thus gained can then also help learners who are not obtaining such good results.

It goes without saying that the complexity of the Swiss educational system makes generalizations difficult; this, however, is a general problem in studies of foreign language learning, which we discuss in detail in Pfenninger and Singleton (forthcoming) and Pfenninger and Singleton (in prep.).

It is also important to bear in mind that in Switzerland, a distinction is made between CLIL and immersion: while activities are undertaken in English in the CLIL classroom, these activities relate to the learning of the second language. As such, the CLIL program in Swiss primary schools 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-minute lessons per week). The emphasis is placed on L2 sensitization, oral fluency, comprehension, cultural awareness, vocabulary and formulaic language. However, the strong focus on meaning in comprehensible input and the communication of authentic messages resemble the

---

<sup>8</sup> Note that our studies show that, naturally, we also find a clear discrepancy between low-proficiency and high-proficiency FL learners as well as more motivated and less motivated students in this population.

main goal of immersion programs. By contrast, the CLIL program that Swiss students later attend in secondary school is a partial immersion program that consists of three content subjects (e.g. 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). Even though in many Swiss schools a student's average school grade functions as a criterion in deciding who can join the program and who cannot, the immersion students in this study did not have significantly better grades in English before they entered the program. This ensures to a certain extent that the results will not be contaminated by the fact that the EARLY CLIL and LATE CLIL groups are more proficient than the EARLY NON-CLIL and LATE NON-CLIL groups to start with.

Finally, it is important to mention that English is considered an L3 here due to the special linguistic landscape in Switzerland: while Swiss German is a High Alemannic variety of German, it is hardly understandable to someone who knows only Standard German, as the two languages differ to some extent in lexicon, phonology and syntax (for a discussion of this, see e.g., Berthele, 2010). According to Lüdi (2007: 161), most Swiss citizens are monolingual during their childhood, but they usually become bilingual in the early primary grades at the latest when they receive formal literacy training in L2 German from 1st grade on (age 7). This means that German-speaking Swiss children have to learn to read, write, and use a relatively unknown language all at once.

### 3.3 MEASURES

Due to the fuzziness of the Ideal L2 Self/Ought-to L2 Self binary in the L2 Motivational Self System proposed by Dörnyei (2005, 2009) as well as the Integrativeness/Instrumentality binary in Gardner's Socio-Educational Model of Language Learning (e.g. Gardner, 2008; Gardner & Lambert, 1959), which emerged in a language experience essay written by the 200 participants (see Pfenninger & Singleton, in prep., for a detailed description of this task), it was decided in this study to make a distinction between learners' *Future selves* and their *Present selves*, rather than between L2 Self, Integrativeness and Instrumentality. *Future selves* encompasses students' wish to become similar to native speakers of English as well as the usefulness of the L2 skills learned in the future. *Present selves* refers to the current attitudes learners display toward EFL and the L2 community and their reactions to a world in

which English plays a predominant role, as well as the extent to which the learners want to be involved in cross-cultural contact situations and travel to English-speaking countries. This dimension also includes those factors of external regulation which lead to action in order to avoid punishment or bad grades or assuage one's guilty conscience. Participants completed a Likert-type questionnaire that consisted of 15 items, which comprised five choices (totally disagree, disagree, neutral, agree, totally agree) for each of the five statements. The 15 items were taken from the motivation questionnaire of a large-scale study (see Pfenninger & Singleton, forthcoming). A third of the statements were formulated in the negative, and the resultant list was translated into German and randomized. Table 1 shows the Cronbach's alpha reliability coefficients for the two multi-item scales of the present study. All of the reliability coefficients are above the recommended .70 threshold.

**Table 1.** Information on the multi-item scales

Variables	No. of items	Cronbach's alpha	Sample item
Future selves	8	.79	Whenever I think of the future, I imagine myself as someone who is able to speak English.
Present selves	7	.75	As a language, I don't like English.

Language data were collected by means of a test battery that included a composition, a grammaticality judgment task,<sup>9</sup> a vocabulary size test (Academic sections in Schmitt, Schmitt and Clapham's (2001) Versions A and B of Nation's Vocabulary Levels Test), the Productive Vocabulary Size Test by Laufer and Nation (1999), and a listening comprehension task (see Pfenninger, 2014). The tasks had been aligned against Level B2/C1 in the Common European Framework of Reference for Languages (CEFR). The grammaticality judgment task included morphosyntactic structures that have been found to be particularly age-sensitive, such as articles and inflections, as well as structures that are not particularly age-sensitive, for instance word order and *do*-support (see, e.g., McDonald, 2006).

<sup>9</sup> The reliability coefficient (KR-20) obtained was .90 for grammatical items and .95 for ungrammatical items.

### 3.4 METHOD

I used R (R Development Core Team 2014) and *lme4* (Bates, Maechler & Bolker, 2008) to perform a linear mixed effects analysis (also called multi-level analysis) of the relationship between AO, CLIL and the L2 Self (see Pfenninger & Singleton, forthcoming, for a discussion of the benefits of such models). As fixed effects, I entered AO, type of instruction, and motivation into the model. Note that when including continuous predictors in a mixed-effect model such as motivation, it is often useful to center each predictor around its mean value (Cunnings, 2012). This involves subtracting from each individual value of a predictor the predictor's overall mean, and is done to help reduce collinearity within the model (e.g. between main effects and interactions; see Jaeger, 2010). The final models had random effects (intercepts) to account for class-to-class and school-to-school differences that induce correlation among scores for students within a school and within a class. In other words, the hierarchical structure of the data on all skills tested consisted of three levels: student (level 1), class (level 2), and school (level 3). The scores on the tests were added to the model at the student level. There were significant random school and class effects for all dependent variables. Likelihood ratio tests showed that random slope models (subject-specific slopes for the fixed effect AO) were not necessary for any dependent measure, so I constructed random intercept models. None of the interactions included (age \* motivation; age \* instruction; instruction \* motivation) provided any better fit, except for one area (productive vocabulary, see below). For the listening comprehension task, the grammaticality judgment task, the productive vocabulary and receptive vocabulary tasks, I added random intercepts for subjects and items in order to account for the fact that some participants may generally have attained higher scores in this particular task than others, and some items may generally have yielded lower scores than others (see Cunnings, 2012: 374).

Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. *P*-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. All models reported were fitted using Laplace estimation with the R software. Also, all models were first evaluated with likelihood ratio tests (test model vs. null model with only the control variables). If the full model vs. null model comparison reached significance, I present *p*-values based on likelihood ratio tests. Given the lack of degrees of freedom with mixed models, I refrain from reporting *df*.

## 4. RESULTS

Table 2 presents the mean scores, standard deviations and intergroup differences for the seven language measures and the motivation measure:

**Table 2.** Descriptive statistics (means and standard deviations)

	EARLY CLIL (n=50)	EARLY NON-CLIL (n=50)	LATE CLIL (n=50)	LATE NON-CLIL (n=50)
LC	18.56 (1.81)	12.7 (3.16)	15.86 (3.16)	12.4 (3.68)
PV	35.52 (7.06)	24.98 (7.16)	35.37 (7.64)	25.65 (7.52)
RV	56.42 (5.27)	49.85 (6.15)	55.44 (6.20)	48.84 (7.84)
W/TU	17.70 (3.43)	14.25 (2.32)	17.10 (4.47)	13.35 (2.57)
CL/TU	1.76 (0.37)	1.58 (0.22)	1.79 (0.30)	1.57 (0.25)
ERR/TU	0.51 (0.29)	0.53 (0.34)	0.50 (0.43)	0.54 (0.30)
GJT	42.83 (2.07)	42.84 (2.84)	42.92 (2.41)	43.10 (2.5)
Motivation (L2 Self)	4.17 (0.63)	3.79 (0.72)	4.18 (0.68)	3.76 (0.66)

*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

To answer the research question regarding the strength of the association between English proficiency with starting age, on the one hand, and with type of instruction and motivation, on the other, mixed linear regression models with the test scores as dependent variables were fitted. A summary of all models is presented in Tables 3 and 4:

**Table 3.** Multilevel regression analyses for the investigated dependent variables (fixed effect estimates)

Fixed effect: AO				
	Coefficient $\beta$	Standard error	$\chi^2$	$p$
LC	-1.25	0.74	3.20	.074
PV	-0.09	1.80	0.02	.879
RV	2.08	1.59	1.26	.261
W/TU	-1.82	2.31	4.87	.199
CL/TU	0.08	0.07	0.78	.379
ERR/TU	0.00	0.05	0.02	.883
GJT	0.53	0.57	0.79	.373

Fixed effect: Instruction				
	Coefficient $\beta$	Standard error	$\chi^2$	$p$
LC	-6.46	2.37	17.43	<.0001**
PV	-5.76	5.89	12.63	<.0001**
RV	-10.32	5.40	14.03	<.0001**
W/TU	-0.63	2.83	10.96	<.0001**
CL/TU	-0.21	0.25	8.13	.004**
ERR/TU	0.21	0.30	0.63	.732
GJT	0.27	2.03	1.62	.203

Fixed effect: Motivation				
	Coefficient $\beta$	Standard error	$\chi^2$	$p$
LC	1.54	0.41	51.43	<.0001**
PV	6.17	1.02	51.03	<.0001**
RV	0.59	0.93	14.91	<.0001**
W/TU	0.96	0.49	11.75	<.0001**
CL/TU	-0.04	0.04	0.57	.449
ERR/TU	0.03	0.05	5.58	.883
GJT	0.42	0.35	20.18	<.0001**

\*Statistically significant at  $\alpha < .05$ ; \*\*Statistically significant at  $\alpha < .01$

*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.

**Table 4.** Interaction between fixed effects

AO * Instruction				
	Coefficient $\beta$	Standard error	$\chi^2$	$p$
LC	2.93	0.93	9.54	.002**
PV	1.02	2.67	0.00	.947
RV	2.48	0.65	1.29	.256
W/TU	0.60	0.34	1.57	.210
CL/TU	0.04	0.10	0.23	.640
ERR/TU	0.03	0.11	0.12	.730
GJT	0.86	0.76	1.06	.304

AO * Motivation				
	Coefficient $\beta$	Standard error	$\chi^2$	$p$
LC	-0.07	0.57	0.01	.930
PV	0.87	1.40	0.77	.380
RV	2.46	1.27	3.63	.067
W/TU	0.41	0.67	0.45	.503
CL/TU	0.07	0.06	1.43	0.23
ERR/TU	0.12	0.07	2.81	.094
GJT	0.46	0.48	1.07	.301

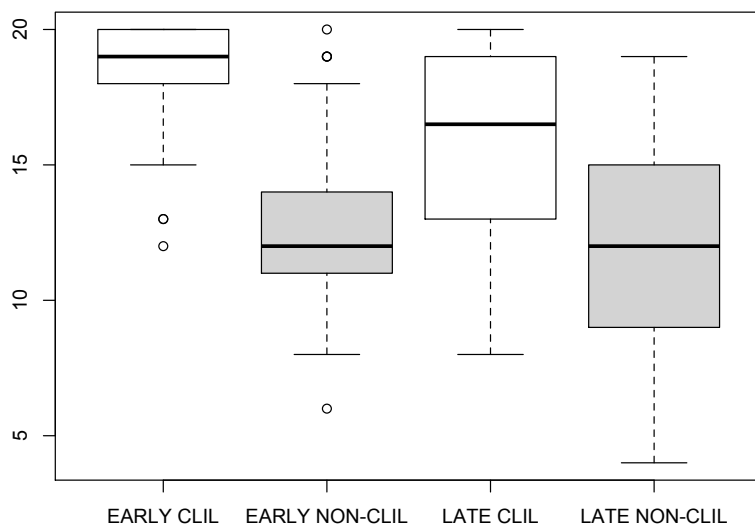
  

Instruction * Motivation				
	Coefficient $\beta$	Standard error	$\chi^2$	$p$
LC	0.62	0.57	1.36	.244
PV	-3.27	1.41	4.20	.040*
RV	1.16	1.30	0.83	.361
W/TU	-0.66	0.69	0.81	.367
CL/TU	0.01	0.06	0.03	.860
ERR/TU	-0.04	0.07	0.47	.491
GJT	0.06	0.49	0.03	.862

\*Statistically significant at  $\alpha < .05$ ; \*\*Statistically significant at  $\alpha < .01$

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.

It is clear from Table 3 that there were no age effects for any of the dependent variables, and AO did not interact with type of instruction or motivation for any of the seven measures, with one exception: there was a significant interaction between AO and type of instruction for listening comprehension, which reflects the advantage of the EARLY CLIL group over all the other groups, as illustrated in Figure 1:

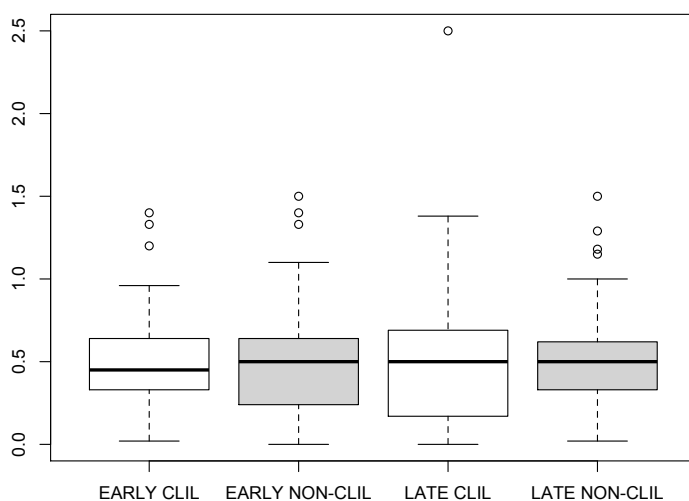


**Figure 1.** Listening scores (LC) by group (actual data points are overlaid on the boxplot and median lines are in bold).

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). Furthermore, older starters did not show greater variation in their L2 performance, as Table 2 above shows (see also Pfenninger 2011, 2014).

CLIL significantly affected five out of seven dependent variables: listening comprehension (increasing it by about  $6.5 \pm 2.4$  points on a 20-point scale), productive vocabulary (raising it by about  $5.8 \pm 5.89$  points on the 54-point scale), receptive vocabulary (increasing it by  $10.32 \pm 5.40$  points on the 60-point scale), fluency (increasing it by  $0.63 \pm 2.83$  words per T-unit), and complexity (enhancing it by  $0.21 \pm 0.25$  per T-unit). Interestingly, both early starters and late starters benefited from immersion with respect to these measures. Accuracy as measured by errors/T-unit and grammaticality judgments was not affected by CLIL (see also Pfenninger, 2014). In fact, accuracy was not affected by either AO, type of instruction or motivation. All four groups had similar scores despite their dissimilar profiles, as Figure 2 shows for productive accuracy:

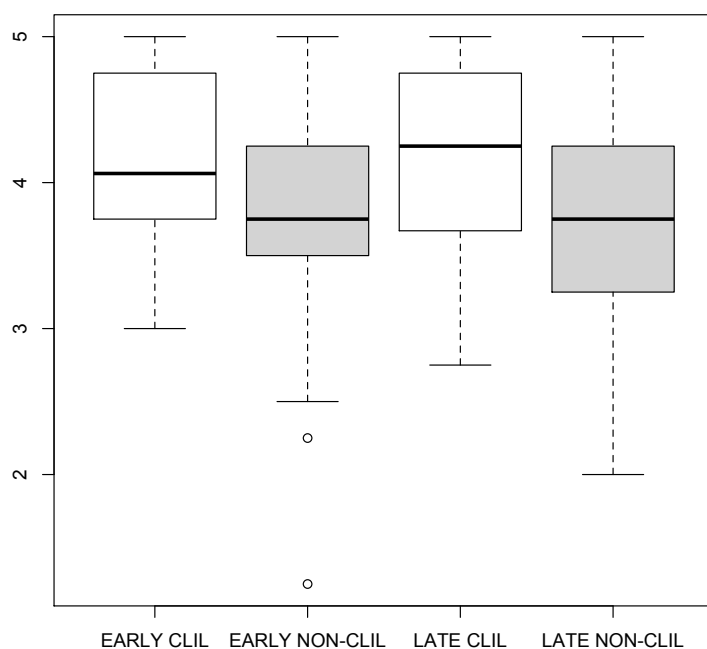




**Figure 2.** Productive accuracy scores (ERR/TU)

Motivation also affected five out of seven dependent variables: listening comprehension (improving it by  $1.54 \pm 0.41$  points on a 20-point scale), productive and receptive vocabulary (increasing them by  $6.17 \pm 1.02$  points and  $0.54 \pm 0.93$  points, respectively), fluency (improving it by  $0.96 \pm 0.49$  words per T-unit), and grammaticality judgments (improving them by  $0.42 \pm 0.35$  points), but it did not have an effect on complexity and accuracy. Interestingly, there was no interaction between motivation and AO or motivation and type of instruction, which indicates that irrespective of starting age or type of instruction received, students with a higher motivation level outperformed less motivated students. The only interaction between motivation and type of instruction was found in the area of productive vocabulary, which was due to the CLIL students' higher motivation.

In order to calculate the effects of AO and type of instruction on motivation, a mixed model was fitted with AO and type of instruction as fixed effects, and school and class as random effects (intercepts). The results showed that whereas AO did not affect motivation ( $\chi^2(1)=0.00$ ,  $p=0.949$ ), instruction type (that is, CLIL in secondary school) had a significant impact ( $\chi^2(1)=12.77$ ,  $p=0.0004$ ), enhancing motivation by about  $0.40 \pm 0.10$  points on a 5-point scale. Figure 4 illustrates the higher motivation of the CLIL students:



**Figure 3.** Motivation by group

Thus, with respect to the second variable under investigation, motivation, the findings suggest the following: the CLIL students were more motivated than their non-CLIL counterparts; however, generally speaking, students with greater motivation performed better on the English tests, irrespective of the type of instruction and AO; as for FL competence, CLIL had more beneficial effects than regular EFL instruction.

## 5. DISCUSSION

The current study has found that late-starting groups (LATE CLIL and LATE NON-CLIL) were able to catch up with the EARLY NON-CLIL group, which supports the hypothesis that the initial fast rate of FL learning of older learners may last for several years in an input-impoverished environment (Larson-Hall, 2008; Muñoz & Singleton, 2011; Singleton, 1995a, 1995b, 2005). Even though early learners (such as the EARLY CLIL and the EARLY NON-CLIL in this study) may in theory have greater potential than late starters due to their earlier AO 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, as we have seen in

the strong performance of the EARLY CLIL with respect to listening skills (see also Cenoz & Jessner, 2009).

Probably more surprising than the EARLY CLIL students outperforming students in non-CLIL programs (EARLY NON-CLIL 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, as well as 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 NON-CLIL and LATE NON-CLIL). The fact that CLIL seems to bear rich fruits with respect to vocabulary has been well documented in the literature (see literature review above). The overall success of the LATE CLIL group in these various skills is yet another indicator that instruction seems capable of overriding the age factor in a classroom setting.

The findings also confirm previous studies (e.g., Collins et al., 2012; Genesee, 1987, 2004; Pica, 2011; Spada & Lightbown, 1989) that found that (morphosyntactic) accuracy remains challenging for CLIL students. They also corroborate the positive effects of form-focused instruction on acquisition, that is, the effectiveness of explicit instruction on students' acquisition and use of specific morphosyntactic features of English. 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 paid great attention to accuracy.

With respect to motivation, the findings confirm previous CLIL research (e.g. Lasagabaster, 2011) suggesting that learning in the FL increases motivation. The novel aspect of this study is that CLIL and motivation had a similar effect on language competence, *without* interaction between them. Finally, the results do not confirm previous findings that the more years students spend studying a subject, the more disenchanted with it they become (see e.g. Davies & Brember, 2001), i.e., AO (and therefore also length of instruction) do not

have a significant effect on motivation. CLIL in secondary school, on the other hand, has a significant impact on students' motivation levels at the end of secondary education.<sup>10</sup>

## 6. CONCLUSION

It was my goal in this study to not see CLIL “in a vacuum”, as Bruton (2011: 531) fears happens in most CLIL studies, but to examine four different real-life educational scenarios that have been or are currently practiced in the Swiss system. As DeKeyser (2012: 190) rightly points out, interactions between individual variables and external, educational or contextual variables allow for more fine-tuned (and hence more generalizable) predictions that help to adapt teaching methodologies to students or curriculum design.

Like so many previous studies (e.g. Freed, Dewey, Segalowitz & Halter, 2004; Lasagabaster, 2011), my analysis has shown that CLIL programs should be boosted as they exert a very positive influence on learners' FL achievement. Since CLIL is not very well established in Switzerland, it still has to struggle for recognition and support. To date, intensive EFL is an optional program available to a minority of (high-achievement) students. In light of the finding that it is particularly low-level learners that make the most impressive progress in an intensive program (see e.g. White & Collins, 2012), it is highly recommended to implement a plan to offer intensive EFL to more secondary school students in Switzerland.

Furthermore, a number of (well-known) problems have emerged in this study, similar to previous studies of the outcomes of CLIL programs:

- (1) One obvious limitation in this study is that since the CLIL groups not only had English 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 provision and exposure (see Bruton, 2011; Cenoz et al., 2014). In other words, the CLIL students received many more hours of (formal and informal) EFL instruction than any of the other groups. This is probably one of the most fundamental issues for CLIL researchers and can only be resolved with complementary qualitative analyses (see Pfenninger & Singleton, in prep.).
- (2) One factor that can be – and has to be – controlled for in the future, however, is aptitude, based on the insight that “CLIL can attract a disproportionately large number of academically bright students” (Mehisto, 2007: 63). It would greatly enrich the CLIL field to analyze the impact of aptitude at the beginning and at the end of

---

<sup>10</sup> In another large-scale study (see Pfenninger and Singleton forthcoming.) I administered end-of-program questionnaires to the students, and the responses consistently show high levels of enthusiasm among all students and considerable confidence in their acquired abilities to express themselves in English.

immersion programs, so that the effect of intensive learning contexts could be more effectively assessed from a theoretical perspective. Even though the four groups in this study started with similar overall academic achievement according to previous grades in English, the CLIL participants might have profited from cognitive advantages that could not be captured in this research design.

- (3) Related to (2), another caveat that needs to be mentioned is that there was no pretest. Even though the four groups in this study started with the same percentage averages in English, the CLIL participants might have profited from cognitive advantages that could not be captured in this research design. Of course, there might also have been language competence differences between the CLIL and non-CLIL groups that were not reflected by the students' grades, as was the case in Alonso, Grisaleña & Campo (2008).
- (4) Because of the diversity of CLIL programs in Europe and the lack of conceptual clarity (see Cenoz et al., 2014), it is difficult for researchers to provide a clear and detailed description of CLIL classrooms/programs.

This calls for further (critical) research into the methodological approach in which foreign language teaching takes place.

In a next step, it also seems to be interesting to analyze which input measures (length of instruction in years, use of English as the language of instruction, number of curricular and extracurricular lessons, amount of time spent in a naturalistic immersion situation abroad, current informal contact with the target language) are more strongly associated with long-term L3 performance and how aptitude factors into this process (see Pfenninger & Singleton, in prep.).

### **Acknowledgments**

The writing of this article was supported by a Forschungskredit of the University of Zurich, Grant FK-15-078. This grant is hereby gratefully acknowledged. I am greatly indebted to the students and teachers for their enthusiastic participation and support. Also, I am grateful to Johanna Gündel for her research assistance, as well as two anonymous reviewers for their helpful comments and suggestions. Any remaining errors remain my own.

**BIBLIOGRAPHICAL REFERENCES**

- ADMIRAAL, Wilfried, Gerard WESTHOFF & Kees DE BOT (2006), Evaluation of bilingual secondary education in The Netherlands: Students' language proficiency in English. *Educational Research and Evaluation* 12 (1), 75-93.
- ALONSO, Esmeralda, Jesus GRISALENA & Alejandro CAMPO (2008). Plurilingual education in secondary schools: analysis of results. *International CLIL Research Journal* 1 (1), 36-49.
- AL-THUBAITI, Kholoud (2010). *Age effects in a minimal input setting on the acquisition of English morphosyntactic and semantic properties by L1 speakers of Arabic*. PhD Thesis, University of Essex, UK.
- BATES, Dough M., Martin MAECHLER & Ben BOLKER (2012). lme4: Linear mixed-effects models using Eigen and Eigen. R package version 0.999999-0. <http://cran.R-project.org/package=lme4>.
- BERTHELE, Raphael (2010). Dialekt als Problem oder Potenzial. Überlegungen zur Hochdeutschoffensive in der deutschen Schweiz aus Sicht der Mehrsprachigkeitsforschung [Dialect: Problem or potential? Reflections on the Standard German debate in German-speaking Switzerland from the perspective of multilingualism research]. In: Bitter Bättig, Franziska and Albert Tanner (eds.), *Sprachen Lernen – Durch Sprache Lernen* [Learning languages – learning through languages]. Zürich: Seismo, 37-52
- BRUTON, Anthony (2011). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System* 39, 523-32.
- CENOZ, Jasone & Ulrike JESSNER (2009). The study of multilingualism in educational contexts. In: Aronin, Larissa and Britta Hufeisen (eds.), *The exploration of multilingualism*. Clevedon: Multilingual Matters, 121-38.
- CENOZ, J., F. GENESEE & D. GORTER (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics* 35 (3), 243-262.
- COLLINS, Laura & Joanna WHITE (2011). An intensive look at intensity and language learning. *TESOL Quarterly* 45 (1), 106-133.
- COLLINS, Laura & Joanna WHITE (2012). Closing the gap: Intensity and proficiency. In C Muñoz (ed.), *Intensive exposure experiences in second language learning*. Bristol: Multilingual Matters, 45-65.

- COLLINS, Laura, Joanna WHITE, Pavel TROFIMOVICH, Walcir CARDOSO & Marlise HORST (2012). When comprehensible input is not comprehensive input: A multi-dimensional analysis of instructional input in intensive English as a Foreign Language. In: Muñoz, Carmen (ed.), *Intensive exposure experiences in second language learning*. Bristol: Multilingual Matters, 66-87.
- COYLE, Do (2008). CLIL – A pedagogical approach from the European perspective. In: Van Dusen-Scholl, Nelleke and Nancy H. Hornberger (eds.), *Encyclopedia of Language and Education*. New York: Springer, 97-111.
- COYLE, Do, Philip HOOD & David MARSH (2010). *CLIL. Content and language integrated learning*. Cambridge: CUP.
- CSIZÉR, Kata & Zoltan DÖRNYEI (2005). The Internal Structure of Language Learning Motivation and Its Relationship with Language Choice and Learning Effort. *The Modern Language Journal* 89 (1), 19-36.
- CSIZÉR, Kata & Gabriella LUKÁCS (2010). The Comparative Analysis of Motivation, Attitudes and Selves: The Case of English and German in Hungary. *System* 38, 1-13.
- CUMMINS, Jim (1995). Canadian French immersion programs: A comparison with Swedish immersion programs in Finland. In: Buss, Martina and Christer Laurén (eds.), *Language immersion: Teaching and second language acquisition from Canada to Europe* (Research No. 192, Linguistics 30). Vaasa: Vaasa University, 7-20.
- CUNNINGGS, Ian (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research* 28, 369-82.
- DALTON-PUFFER Christiane (2007). *Discourse in Content and Language Integrated Learning (CLIL) Classrooms*. Amsterdam and Philadelphia: John Benjamins.
- DALTON-PUFFER, Christiane, Tarja NIKULA & Ute SMIT (2010). Language Use and Language Learning in CLIL: Current Findings and Contentious Issues. In: Dalton-Puffer, Christiane, Tarja Nikula and Ute Smit (eds.), *Language Use and Language Learning in CLIL Classrooms*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- DAVIES, Julie & Ivy BREMBER (2001). The closing gap in attitudes between boys and girls: A five-year longitudinal study. *Educational Psychology* 21 (1), 103-115.
- DE GRAAFF, Rick & Alex HOUSEN (2009). Investigating effects and effectiveness of L2 instruction. In: Long, Mike and Catherine Doughty (eds.) *The Handbook of Language Teaching*. Oxford: Wiley-Blackwell, 726-753.

- DEKEYSER, Robert M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22 (4), 499-533.
- DEKEYSER, Robert M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning* 62 (Suppl. 2), 189-200.
- DEKEYSER, Robert M. & J. LARSON-HALL (2005). What does the critical period really mean?. In: Kroll, Judith F., and Annette M. B. de Groot (eds.), *Handbook of Bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press, 89-108.
- DÖRNYEI, Zoltán (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- DÖRNYEI, Zoltán (2009). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning* 59 (Suppl. 1), 230-248.
- DÖRNYEI, Zoltán & Letty CHAN (2013). Motivation and Vision: An Analysis of Future L2 Self Images, Sensory Styles, and Imagery Capacity Across Two Target Languages. *Language Learning* 63 (3), 437-462.
- ELLIS, Nick C (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.
- ELLIS, Rod (2004). Individual differences in second language learning. In: Davies, A. and C. Elder (eds.) *The Handbook of Applied Linguistics*. Oxford: Blackwell Publishing, 525-551.
- Englisch an der Primarschule. Grundsatz und Rahmenbedingungen. Erwägungen zum Bildungsratsbeschluss und Bildungsratsbeschluss vom 18.3.2003* (2003), Zürich: Bildungsrat. Available at <http://edudoc.ch/record/28911/files/40.pdf>.
- EUROBAROMETER 243 (2006). Europeans and their languages. [http://ec.europa.eu/education/languages/pdf/doc629\\_en.pdf](http://ec.europa.eu/education/languages/pdf/doc629_en.pdf) (15 July 2008).
- EUROPEAN COMMISSION (1995). *White Paper on Education and Training - Teaching and Learning - Towards the Learning Society*. <[http://ec.europa.eu/white-papers/index\\_en.htm#block\\_13](http://ec.europa.eu/white-papers/index_en.htm#block_13)> (accessed 08 January 2015).
- FLEGE, James E & Serena LIU (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23 (4), 527-552.
- FREED, Barbara F., Dan P. DEWEY, Norman SEGALOWITZ & Randall HALTER (2004). The language contact profile. *Studies in Second Language Acquisition* 26 (2), 349-56.



- GARCÍA MAYO, María del Pilar & María L GARCÍA LECUMBERRI (2003). *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.
- GARDNER, Robert C. (1985). *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. London: Edward Arnold.
- GARDNER, Robert C. & Wallace E. LAMBERT (1959). Motivational variables in second language acquisition. *Canadian Journal of Psychology* 13, 266-272.
- GENESE, Fred (1987). *Learning through Two Languages: Studies of Immersion and Bilingual Education*. Cambridge, MA: Newbury House.
- GENESE, Fred (2004). What Do We Know About Bilingual Education For Majority Language Students?. In: Bhatia, Tej K. and William Ritchie (eds.), *Handbook of Bilingualism and Multiculturalism*. Malden, MA: Blackwell, 547-576.
- HAENNI HOTI, Andrea U., Sybille HEINZMANN, Marianne MÜLLER, Werner WICKI, Marta OLIVEIRA & Erika WERLEN (2011). Introducing a second foreign language in Swiss primary schools: The effect of L2 listening and reading skills on L3 acquisition. *International Journal of Multilingualism* 8, 98-116.
- HULSTIJN, Jan (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research* 18 (3), 193-223.
- JAEGER, Florian (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language* 59 (4), 434-446.
- KINSELLA, Ciara & David SINGLETON (2014). Much more than age. *Applied Linguistics* 35 (4), 441-462.
- KORMOS, Judit & Kata CSIZÉR (2008). Age-related Differences in the Motivation of Learning English as a Foreign Language: Attitudes, Selves and Motivated Learning Behaviour. *Language Learning* 58 (2), 327-355.
- LARSON-HALL, Janet (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research* 24 (1), 35-63.
- LASAGABASTER, David (2011). English achievement and student motivation in CLIL and EFL settings. *Innovation in Language Learning and Teaching* 5 (1), 3-18.
- LAUFER, Batia & Paul NATION (1999). A vocabulary size test of controlled productive ability. *Language Testing* 16 (1), 33-51.

- LONG, Mike H (1981). Input, interaction and second language acquisition. *Annals of the New York Academy of Sciences* 379, 259-78.
- LÜDI, Georges (2007). Sprachverhalten, Sprachpolitik, Diskurs über Sprache: Staatlichkeit in Europa zwischen dem einsprachigen Nationalstaat und dem mehrsprachigen Vielvölkerstaat [Linguistic behavior, language policies and public discourse about language: Statehood in Europe between the monolingual nation state and the multilingual multiethnic state]. In: Nekula, Marek (ed.), *Sprache und nationale Identität in öffentlichen Institutionen der Kafka-Zeit* [English and national identity in public institutions in Kafka's time]. Köln: Böhlau, 13-30.
- MCDONALD, Janet 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.
- MAILLAT, Didier (2010). The pragmatics of L2 in CLIL. In: Dalton-Puffer, Christiane, Tarja Nikula and Ute Smit (eds.), *Language Use and Language Learning in CLIL Classrooms*. Amsterdam and Philadelphia: John Benjamins, 39-58.
- MARSH, David (ed.) (2002). *CLIL/EMILE the European Dimension*. Jyväskylä: University of Jyväskylä.
- MEHISTO, Peeter (2007). What a school needs to consider before launching a CLIL program: the Estonian experience. In: Marsh, David and Dieter Wolff (eds.), *Diverse contexts – converging goals: CLIL in Europe*. Frankfurt: Peter Lang, 61-77.
- MEHISTO, Peeter (2008). CLIL counterweights: Recognising and decreasing disjuncture in CLIL. *International CLIL Research Journal* 1, 93-119.
- MERCER, Sarah & Marion WILLIAMS (eds.) (2014). *Multiple Perspectives on the Self in SLA*. Bristol: Multilingual Matters.
- MIHALJEVIC, DJIGUNOVIC Jelena (2014). L2 learner age from a contextualised perspective. *Studies in Second Language Learning and Teaching* 4 (3), 419-441.
- MOYER, Alene (2014). What's age got to do with it? Accounting for individual factors in second language accent. *Studies in Second Language Learning and Teaching* 4 (3), 443-464.
- MUÑOZ, Carmen (2006). The effects of age on foreign language learning: the BAF project. In: Muñoz, Carmen (Ed.), *Age And The Rate Of Foreign Language Learning*. Clevedon: Multilingual Matters, 1-40.

- MUÑOZ, Carmen (2011). Is input more significant than starting age in foreign language acquisition?. *International Review of Applied Linguistics* 49(2), 113-133.
- MUÑOZ, Carmen (Ed.) (2012a). *Intensive Exposure Experiences In Second Language Learning*. Bristol: Multilingual Matters.
- MUÑOZ, Carmen (2012b). The significance of intensive exposure as a turning point in learners' histories. In: Muñoz, Carmen (ed.), *Intensive Exposure Experiences in Second Language Learning*. Bristol: Multilingual Matters, 141-160.
- MUÑOZ, Carmen (2014). Starting age and other influential factors: insights from learner interviews. *Studies in Second Language Learning and Teaching* 4 (3), 465-484.
- MUÑOZ Carmen & David SINGLETON (2007). Foreign accent in advanced learners: two successful profiles. *The EUROSLA Yearbook* 7, 171-90.
- MUÑOZ, Carmen & David SINGLETON (2011). A Critical Review of Age-related Research on L2 Ultimate Attainment. *Language Teaching* 44 (1), 1-35.
- MYLES Florence & Rosamond MITCHELL (2012). Learning French from ages 5, 7 and 11: An investigation into starting ages, rates and routes of learning amongst early foreign language learner. *ESRC End of Award Report* RES-062-23-1545.
- NETTEN, Joan & Claude GERMAIN (2004). Theoretical and research foundations of Intensive French. *Canadian Modern Language Review* 60 (3), 275-294.
- NIKOLOV, Marianne and Jelena MIHALJEVIC DJIGUNOVIC (2011). All shades of every color: An overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics* 31, 95-119.
- PÉREZ-CANADO, Maria L. (2012). CLIL research in Europe: past, present, and Future. *International Journal of Bilingual Education and Bilingualism* 15 (3), 315-341.
- PÉREZ-VIDAL, Carmen & María JUAN-GARAU (2010). To CLIL or not to CLIL: From bilingualism to multilingualism in Catalan/Spanish communities. In: Lasagabaster, David and Yolanda Ruiz de Zarobe (eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. Newcastle-Upon-Tyne: Cambridge Scholars Publishing, 115-138.
- PFENNINGER, Simone 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.
- PFENNINGER, Simone E. (2014). The misunderstood variable: Age effects as a function of type of instruction. *Studies in Second Language Learning and Teaching* 4 (3), 529-56.

- PFENNINGER, Simone E. & David SINGLETON (forthcoming). Recent advances in quantitative methods in age-related research. In: Navracscics, Judit and Simone E. Pfenninger (eds.), *Implications for the Future: Perspectives from Applied Linguistics*. Bristol: Multilingual Matters.
- PFENNINGER, Simone E. & David SINGLETON (submitted). A person-in-context relational view of age and motivation. *Second Language Research*.
- PFENNINGER, Simone E. & David SINGLETON (in prep.). *Beyond Age Effects: Facets, Facts And Factors Of Foreign Language Instruction In A Multilingual State*. Bristol: Multilingual Matters.
- PICA, Teresa (2011). Second Language Acquisition Research: Applied and Applicable Orientations to Practical Questions and Concerns. In: Hinkel, Eli (ed.) *Handbook of Research in Second Language Teaching and Learning*, Volume II. London: Routledge, 257-273.
- R Core Team (2012). *R: A Language And Environment For Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- RUIZ DE ZAROBÉ, Yolanda & Rosa M. JIMÉNEZ CATALÁN (2009). *Content and Language Integrated Learning: Evidence from Research in Europe*. Clevedon: Multilingual Matters.
- SCHMITT, Norbert, Diane SCHMITT & Caroline CLAPHAM (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test, *Language Testing* 18 (1), 55-88.
- SERRANO, Raquel & Carmen MUNOZ (2007). Same hours, different time distribution: Any difference in EFL?. *System* 35 (3), 305-321.
- SEIKKULA-LEINO Joana (2007). CLIL learning: Achievement level and affective factors. *Language and Education* 21 (4), 328-341.
- SINGLETON, David (1995a). Introduction: A critical look at the Critical Period Hypothesis in second language acquisition research. In: Singleton, David and Z. Lengyel (eds.), *The Age Factor In Second Language Acquisition: A Critical Look At The Critical Period Hypothesis*. Clevedon: Multilingual Matters, 1-29.
- SINGLETON, David (1995b). Second languages in the primary school: The age factor dimension. *Teanga: The Irish Yearbook of Applied Linguistics* 15, 155-166.
- SINGLETON, David (2005). The Critical Period Hypothesis: A coat of many colours. *International Review of Applied Linguistics in Language Teaching* 43 (4), 269-285.

- SPADA, Nina & Patsy M. LIGHTBOWN (1989). *Intensive ESL programs in Quebec primary schools*. TESL Canada Journal 7, 11-32.
- SWAIN, Merrill (1985). Bilingual education for the English-speaking Canadian. In: Alatis, James E. and John Staczek (eds.), *Perspectives On Bilingualism And Bilingual Education*, Washington, DC, Georgetown University Press, 385-398.
- Swiss Conference of Cantonal Ministers of Education (EDK) (2014). Fremdsprachen: Sprache, Beginn,  
<[http://www.edudoc.ch/static/strukturdaten/pdf\\_rohdaten/092.pdf](http://www.edudoc.ch/static/strukturdaten/pdf_rohdaten/092.pdf)> and  
<[http://www.edudoc.ch/static/strukturdaten/pdf\\_rohdaten/094.pdf](http://www.edudoc.ch/static/strukturdaten/pdf_rohdaten/094.pdf)> (accessed 06 Jan 2015).
- Swiss Conference of Cantonal Ministers of Education (EDK) (2015), *Faktenblatt – Fremdsprachenunterricht in der obligatorischen Schule*. Berne: EDK.  
<[http://www.edudoc.ch/static/web/arbeiten/sprach\\_unterr/fktbl\\_sprachen\\_d.pdf](http://www.edudoc.ch/static/web/arbeiten/sprach_unterr/fktbl_sprachen_d.pdf)>  
(accessed 16 April 2015).
- UNSWORTH, Sharon, Kees DE BOT, Liv PERSSON & Tineke PRINS (2012). *Foreign Languages in Primary School Project. Proceedings of the Foreign Languages in Primary Schools Projects: presentation results FLiPP-research*. Amersfoort, 13 December 2012.
- VESTERBACKA, Selon (1991). *Elever i Språkbudskola, Social Bakgrund och Tidig Språkutveckling* [Summary, children in immersion programme, home background and early language development: Issues and strategies]. Newbury Park, CA: Sage.
- WINITZ, Harris, Brenda GILLESPIE & Jennifer STARCEV (1995). The development of English speech patterns of a 7-year-old Polish-speaking child. *Journal of Psycholinguistic Research* 24 (2), 117-143.