



Studies in Second Language Learning and Teaching

Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz

SSLT 7 (1). 2017. 19-46

doi: 10.14746/sslit.2017.7.1.2

<http://www.sslt.amu.edu.pl>

Not so individual after all: An ecological approach to age as an individual difference variable in a classroom

Simone E. Pfenninger

University of Salzburg, Switzerland

simone.pfenninger@sbg.ac.at

Abstract

The main goal of this paper is to analyze how the age factor behaves as an alleged individual difference (ID) variable in SLA by focusing on the influence that the learning context exerts on the dynamics of age of onset (AO). The results of several long-term classroom studies on age effects will be presented, in which I have empirically analyzed whether AO works similarly across settings and learners or whether it is influenced by characteristics of the setting and the learner—and if so, whether there are contextual variables that can help us understand why those outcomes are different. Results of multilevel analyses indicate that macro-contextual factors (i.e., the wider school context) turn out to have a mediating effect on the relation between AO and L2 proficiency increase, exerting both positive and negative influences and thus suggesting that AO effects are malleable, which is what one would expect if we are dealing with an ID variable. In contrast, no such phenomenon can be observed in relation to lower contextual levels; learners within *classes* do not vary with regard to how sensitive they are to AO. Since the broader social environment in which learning takes place seems to be more influential than the cognitive state assumed to be a characteristic of the individual, I suggest that an ID model that assumes that age is a “fixed factor” (Ellis, 1994, p. 35) is not entirely satisfactory.

Keywords: age factor; context; environmental variables; young learners; individual differences

1. Introduction

Age is often discussed as if it were a simple, single factor that is “beyond external control” (Ellis, 1994, p. 35). This is despite the fact that for many years it has been authoritatively pointed out that ignoring context when it comes to understanding individual differences (IDs) between learners leads to a spurious, or, at least incomplete understanding; as Larsen-Freeman (2015, p. 16) poignantly puts it, “with the coupling of the learner and the learning environment, neither the learner nor the environment is seen as independent, and the environment is not seen as background to the main developmental drama.” Although it is statistically possible to separate the learner from context, it is untenable to do so because this would carry the implication that the two are independent (van Geert & Steenbeek, 2008).

In this paper, I focus on school contexts that can exert a facilitative, neutral, or inhibitory influence on age of onset (AO). The results of longitudinal and cross-sectional studies on effects of AO are presented, in which I have analyzed whether different schools, classes and participants vary with regard to how sensitive they are to the manipulation at hand (i.e., AO). In a first step, it is essential to test whether AO works similarly across broader school contexts, or whether it is influenced by characteristics of the context—and if so, whether there are macro-contextual variables that can help us understand why those outcomes are different. In a further step, I analyze whether effects of age are different for subjects in different classes and thus subject to micro-contextual variables.

As we will see, the characteristics of the groups under investigation have implications not only for theoretical discussions of the age factor but also for methodology in age-related research. A multilevel modeling approach was deployed to shed light on the way in which AO interacts with macro-contextual variables such as school effects or treatment variables (e.g., type of instruction) and micro-contextual variables such as classroom and clustering effects. I will argue that the use of multilevel models enables us to integrate individual-level and contextual-level data in order to assess the impact of context-varying factors in relation to ID variables. The data suggest that, owing to its complex status as a “macro-variable” co-varying with environmental factors (Montrul, 2008, p. 1), the question of age as an ID variable warrants an entirely separate treatment from most other IDs (but see de Bot & Fang, this issue).

2. Age as an individual difference variable

The usual line is to place age alongside ID variables like gender, aptitude, motivation, learning styles, learning strategies and personality (see e.g., DeKeyser,

2012; Paradis, 2011; Robinson, 2002; Zafar & Meenakshi, 2012). In his seminal overview of individual learner differences, Dörnyei (2005, p. 4) defines IDs broadly as “enduring personal characteristics that are assumed to apply to everybody and on which people differ by degree.” According to Ellis (2006), the study of IDs in SLA research seeks answers to four basic questions:

1. In what ways do language learners differ?

Chronological/biological age and initial age of learning (or age of onset; AO) both have an impact on the affective and linguistic development of learners. While it has been argued that both may impact on L2 achievement by confounding with cognitive factors, education, and other background variables (Bialystok & Hakuta, 1999), several scholars (e.g., Muñoz, 2008) have made the case for the claim that a confound between chronological age and AO may partly explain the negative effect on the performance of the youngest learners in comparison with older learners in school settings, and may thus contribute to the positive relationship between L2 proficiency and older age of learning (see also Question 3 below). Referring to Stevens (2006), Muñoz (2008) points out that chronological age is not just an indicator of biological processes associated with senescence; it is also an excellent indicator of life-cycle stage, strongly associated with motivations and opportunities to speak and to maintain or improve proficiency in an L2.

2. What effects do these differences have on learning outcomes?

Depending on the setting, an earlier AO might lead to better outcomes; for instance, in naturalistic settings age is widely recognized as a robust predictor of long-term success in second language acquisition (cf. Hyltenstam, 1992; Johnson & Newport, 1989; Krashen, Long, & Scarcella, 1979; Patkowski, 1980; Snow & Hoefnagel-Höhle, 1978). However, it is not the case that *everyone* who begins learning an L2 in childhood in an informal setting ends up with a perfect command of the language in question; nor is it the case that those naturalistic learners who begin the L2 later in life inevitably fail to attain the levels reached by younger beginners (see e.g., Kinsella & Singleton, 2014). Related to this, attempts to define the temporal boundaries of a so-called critical or sensitive period for SLA and FL learning have failed, that is, led to inconclusive results as it has not been possible to confidently establish the existence of a discontinuity in the age of arrival/ultimate attainment function (see e.g., Vanhove, 2013).

Furthermore, generalizing age-related outcomes found in naturalistic settings to other contexts, notably the very different context of the classroom, has not been upheld by classroom research. Numerous classroom studies throughout

the world (see e.g., Al-Thubaiti, 2010 for Saudi Arabia; Muñoz, 2006, 2011 for Catalonia/Spain; Larson-Hall, 2008 for Japan; Myles & Mitchell, 2012 for Great Britain; Pfenninger, 2013, 2014a, 2014b for Switzerland; Unsworth, de Bot, Persson, & Prins, 2012 for the Netherlands) have presented consistent results that there are very few linguistic and extra-linguistic advantages to beginning the study of a FL earlier in a minimal-input situation.

Finally, there seem to be different windows for different language domains. In many naturalistic studies (e.g., Clahsen & Felser, 2006; DeKeyser, Alfi-Shabtay, & Ravid, 2010; Granena & Long, 2013; McDonald, 2006, 2008), it is pointed out that L2 morpho-syntax seems to be more vulnerable to processing difficulties than L2 lexico-semantics and more susceptible to age. Such difficulties have been linked to resource limitations that might lead to the inability (a) to access and retrieve stored L2 knowledge (semantically-related difficulties) and/or (b) to detect phonological discriminations in the input (phonologically-related difficulties), similar to the difficulties of native speakers under specific types of stress manipulation (McDonald & Roussel, 2010; Pfenninger, 2011).

3. How do learner differences affect the process of L2 acquisition?

Although the prognosis for the level of “ultimate L2 attainment” (if there ever is such a state) generally deteriorates with increasing AO in a naturalistic setting, older children and adults often proceed faster through early stages in the acquisition of L2 morphology and syntax, that is, they profit from a rate advantage (e.g., García Lecumberri & Gallardo, 2003; Muñoz, 2006; Singleton & Ryan, 2004). Not only is there no evidence that an early start in foreign language learning leads to higher proficiency levels after the same amount of instructional time, but the “jump start” that older learners experience often enables them to catch up relatively quickly with the performance of earlier starters (see e.g., Pfenninger, 2011; Pfenninger & Singleton, 2017) so that younger starters with more instructional time have often failed to show a particularly substantial advantage in terms of long-term proficiency benefits.

4. How do individual learner factors interact with instruction in determining learning outcomes?

DeKeyser (2012) discusses age-by-treatment interaction research in the narrow sense, suggesting that different learning processes are at work at different ages, which may imply the need for different treatment (implicit instruction for younger students vs. traditional teaching methodology for older students). Sze (1994) mentions that since classroom-based L2 learning is generally more cognitively

oriented than naturalistic acquisition, there is more reason to believe that the older instructed learner, whose cognitive ability is more developed, will outperform the younger learner in the L2 classroom. Lightbown (2003, p. 8) points out that, "in instructional settings where the total amount of time is limited, instruction may be more effective when learners have reached an age at which they can make use of a variety of learning strategies, including their L1 literacy skills, to make the most of that time." It is important to note that the "older-is-better" trend has also been found in partial and full immersion programs (see e.g., Genesee, 1987; Harley, 1986). For instance, in some of my earlier studies (Pfenninger, 2014a, 2016), learners who experienced intensive exposure to EFL in late immersion presented similar levels of proficiency in the FL to children who had experienced more exposure to the FL in early immersion programs.

Despite our increasing knowledge of the age factor in SLA, there are still many points that are not understood very well. For instance, there is less agreement about the *reasons* for age effects and the mediating effect of cognitive, affective and environmental factors on age effects (Granena & Long, 2013). Are children at an advantage for neurological or neuro-cognitive reasons (effects of aging on L2 learning) or because of age-related circumstances and contextual factors (e.g., positive attitudes, open-mindedness, greater commitment of time and/or energy, support system, school environment, etc.)? Furthermore, owing to the complexity of the age factor, the question has arisen in recent years if this variable should really be regarded as an ID variable. Ellis (1994) belongs among the few scholars who exclude age from the inventory of IDs. He takes the view that age transcends these categories and potentially impacts on all four, thus *contributing* to, rather than representing, IDs in L2 learning. On the other hand, he considers age to be an example of a "fixed factor" or "general factor," in the sense that "it is beyond external control" (p. 35). By contrast, motivation is for him an example of a factor that is variable and malleable as the strength of an individual learner's motivation can change over time and is influenced by external factors. AO can also be causative (i.e., have an effect on learning as well as on other IDs such as motivation), yet, of course, it cannot be resultative (i.e., be influenced by learning). While it is certainly true that no treatment could alter someone's AO, and the impact of an early or late AO does not change with time, age *effects* are sensitive to, and thus mediated by, contexts and situations, as I will illustrate in this paper.

3. Quo vadis: Taking a person-in-context relational view of age

In general, research on IDs has primarily focused on examining individual learners' cognitive and affective states in relation to goals, intentions, and self-images

and how these factors differ across individuals (Dörnyei, 2005; Kozaki & Ross, 2011). However, we know that ID variables often interact with external variables, thus creating a joint impact on the outcome variable. Hence, in order to complete the “individual differences” model of age outlined above, which assumes that AO is a fixed factor and the individual learner is the epicenter of cognitive processes that drive successful language learning, external factors need to be addressed as environmental influences (e.g., the impact of the learning context or compositional effects within the sample) that impact on and possibly mediate age effects in that age effects disappear as soon as external factors are taken into account—hence an “ecological” or, to use Ushioda’s (2008) words, “person-in-context-relational” view of age.

3.1. Macro-contextual variation: School effects

One of the central goals of applied linguistics has been to place questions of language in their social context, as learners are influenced by context and they in turn help shape the context itself as time progresses (de Bot, Lowie, & Verspoor, 2007; Larsen-Freeman & Cameron, 2008). In motivation research, “macro-contextual variation,” that is, variation in motivation that is driven by broader outside effects such as societal and cultural influences, has been well-documented. Dörnyei (2005, p. 67) holds that, unlike other school subjects, learning a foreign language can be heavily affected by socio-cultural factors “such as language attitudes, cultural stereotypes, and even geopolitical considerations.”

In the same way, AO does not work similarly across settings, that is, it is influenced by characteristics of the setting. As mentioned above, there is good supportive evidence that under certain optimal learning circumstances (that is, high quality, quantity and intensity of input in a naturalistic setting, ample opportunities for interaction with a variety of native speakers, high motivation, etc.), an early AO can indeed explain why some learners succeed more than others.

Similarly, in a school context, school effects indicate the relationship between school characteristics and learning outcomes. School characteristics comprise context variables (e.g., school location, resources, school socioeconomic composition, teacher education and experience), which are beyond the direct control of parents, teachers and administrators, and climate variables (e.g., administrative policies, instructional organization, school operation, values, and expectations of students, parents and teachers) (Ma, Ma, & Bradley, 2008). School-effects research has consistently shown that school policies and practices not only vary in their schooling outcomes, but that they can also improve the levels of schooling outcomes and reduce inequalities between different groups (e.g., lowering high status and/or lifting low status groups). Thus, while

students bring into their schools different individual and family characteristics (see e.g., Haenni Hoti & Heinzmann, 2012) as well as different cognitive and affective conditions, schools are seen to channel or process, through school context and school climate, students with different backgrounds.

3.2. Micro-contextual variation: The complexity of the classroom context

Let us now zoom in on the micro-context, that is, micro-contextual variation due to classroom effects. Language learning in the classroom context has long been recognized as a complex dynamic system in that “individuals are intrinsically joined to their environment and context does not therefore represent a static external variable but is in reality part of the individual” (King, 2015, p. 1). Under classroom effects we understand a complex interplay between effects of individual characteristics including self-confidence, personality, emotion, motivation, degrees of learners’ control over their learning, perceived opportunity to communicate and willingness to communicate, and classroom environmental conditions such as topic, task, interlocutor, receptivity to the teacher and pedagogical approach, and classroom dynamics (see e.g., Borg, 2006; Cao, 2011; Wen & Clément, 2003). It is Chaudron’s (2001) view that classroom processes are heavily influenced by the structure of classroom organization, in which different patterns of teacher-student interaction, group work, degrees of learners’ control over their learning, and variations in tasks and their sequencing, play a significant role in the quantity and quality of learners’ production and interaction with the target language. Another important component of the classroom atmosphere is group size. Cao (2011, p. 472) suggests that “generally students prefer small group or pair work to whole-class activity in both ESL and EFL settings.” Smaller classes may also facilitate more peer communication and mutual understanding, as Dewaele and MacIntyre (2014, p. 264) point out: “Smaller groups are more conducive to closer social bonds, a positive informal atmosphere, and to more frequent use of the FL.” Additional factors include teacher characteristics, which are likely to also raise or lower the outcome for a given classroom (e.g., Borg, 2006). It is unavoidable that the teacher plays an influential role in affecting students’ engagement (see also Cao 2011; Wen & Clément, 2003).

As a consequence of classroom effects, learners can exert a normalizing influence in FL classrooms that can augment or undermine individual learners’ own motivations to learn the FL (see Pfenninger & Singleton, 2016). As early as 1988, van Lier described the importance of taking such classroom effects into consideration in classroom research:

At some point all these factors must be taken into account, for all are relevant, many are related, and as yet we know little about their potential contribution to L2 language development . . . It is clear that, unless we are to oversimplify dangerously what goes on in classrooms, we must look at it from different angles, describe accurately and painstakingly, relate without generalizing too soon, and above all not lose track of the global view, the multifaceted nature of classroom work. (p. 8)

I will argue in this paper that it is not enough for researchers to merely draw connections between language and context, but context needs to be granted appropriate weight in the analyses. Although cohort effects have been observed in age research, too (see e.g., Moyer, 2014; Muñoz, 2014; Nikolov, 2009, p. 93), more often than not, observations of such effects are neglected in the methodological analyses. Indeed, many applied linguists (see e.g., Pennycook, 2005, p. 796) caution that one of the shortcomings of work in applied linguistics generally has been a tendency to operate within “decontextualized contexts.”

4. The study

4.1. Research questions

The following research questions are addressed in this paper:

- RQ1. To what extent is AO mediated by classroom effects?
- RQ2. Can we find some external (e.g., class-level) variables that explain between-group differences more accurately than age effects?
- RQ3. Is the effect of AO different for different schools, classes, tasks and subjects?

Studying interactions between age and external, educational or contextual variables is important as it allows for more fine-tuned (and hence more generalizable) predictions that help with adaptation of teaching methodologies to students or matching students with treatments (understood here as any kind of educational intervention at any level of generality, such as curriculum design, teaching method, content presentation, or practice activity; see DeKeyser, 2012, p. 190).

4.2. Participants

One part of the study has a longitudinal design comprising a random sample of two groups following two different educational models of FL learning in the canton of Zurich ($N = 200$). 100 of them were so-called “early classroom learners” (henceforth ECLs); they were schooled according to the new model and learned Standard German from the first grade onwards, English from the third grade onwards and French from the fifth grade onwards, while 100 were “late classroom learners”

(LCLs), schooled according to the old system without English instruction at primary level (AO 13, Year 7), learning only Standard German from the first grade and French from the fifth grade onwards. The average self-reported age of the students was 13.6 years at the first data collection time (at the beginning of secondary school) and 18.8 years at the second measurement briefly before graduation.

For the qualitative analysis, I selected a focus group of 20 early learners and 20 late learners from those 200 who had participated in the quantitative phase. Early and late learners were selected according to scores on a range of L2 proficiency tests administered at Times 1 and 2. Following Muñoz (2014), the criterion for inclusion in the high achievement groups was a score in the 75th percentile on all tasks, and for inclusion in the low achievement groups a score in the 25th percentile on all tests. Furthermore, the high-achievers all had grades at or above 5 (6 being the highest grade). Following these grouping criteria, I ended up with four groups of 10 participants: 10 early learners, high achievement (ELH); 10 early learners, low achievement (ELL); 10 late learners, high achievement (LLH); and 10 late learners, low achievement (LLL). This focus group was chosen so as to get a better, more detailed impression of students' language learning experiences and beliefs (see below).

Finally, a third group of participants was recruited in the canton of Schaffhausen, where the Early English program is conducted during four years of primary school, that is, the ECLs' AO is around 9 years, whereas the LCLs from the previous curriculum started their English instruction at the age of 13. During a phase of transition, some of the ECLs and LCLs were integrated in the same classes when they entered the academically oriented high school (at around age 15), which provided me with a sample of five mixed classes ($N = 98$; 51 ECLs, 47 LCLs) to investigate class-specific slopes (the effect of AO for different tasks and subjects). The participants were in Grade 9 (mean age: 15.1, range 14-17).

4.3. Procedure

Language data were collected by means of a test battery that included a standardized listening comprehension task, two written compositions (an argumentative and a narrative essay), a grammaticality judgment task,¹ a vocabulary size test (Academic sections in Schmitt, Schmitt, and Clapham's [2001] Versions A and B of Nation's Vocabulary Levels Test), Laufer and Nation's (1999) Productive Vocabulary Size Test, and two oral tasks (the re-telling of a silent movie and a spot-the-difference task) (for a description of these, see Pfenninger & Singleton, 2017). In order to give a better account of the interaction of AO and other (often

¹ The reliability coefficient (KR-20) obtained was .90 for grammatical items and .95 for ungrammatical items.

hidden) variables such as motivation, attitudes and beliefs, the participants were given 45 minutes to write language experience essays, which I hoped would elicit (a) the participants' reflections on their experience of early or late FL learning at the beginning and at the end of secondary school, (b) the participants' affect in respect of foreign languages, and English in particular, and (c) participants' beliefs about the age factor. Loose guidelines were provided for the writing. No specific length was set; students wrote between 203 and 475 words (see Pfenninger & Singleton, 2016).

4.4. Method

The main question is how to operationalize an ecological perspective of the age factor in different settings as described above, for example, the interrelationship between starting age and macro-contextual variables such as school effects or treatment variables (e.g., type of instruction), as well as micro-contextual variables such as classroom and clustering effects. The most frequently used statistics in SLA—general linear models (GLMs) that compare means as a default, as well as correlation-type statistics (e.g., Plonsky, 2013, 2014; Plonsky & Gass, 2011)—are not suitable for a nuanced account of exactly what goes on in the classroom as they run on the averaged data and thus cannot directly provide information about individual change or capture the complexity of contextual effects on individual learning. Besides the problem of the loss of information in GLM, these models are often used in violation of at least some of the assumptions of the procedure, such as the inclusion of correlated errors in linear models. Performance as well as affective factors correlate between the members of one cluster, resulting in the loss of independence among observations, a serious violation of a key assumption underlying a large majority of parametric statistics procedures (e.g., Goldstein, 1995; Raudenbush & Bryk, 2002).

Multilevel modeling (MLM), a subgroup of linear mixed-effects regression modeling, has for some time finally been finding its way into certain SLA subfields (see Pfenninger & Singleton, 2017). The use of multilevel models enables us to integrate individual-level and contextual-level data in order to assess the impact of context-varying factors in relation to ID variables. MLM can also take account of the fact that performance correlates between students within the same class (and school) in a way that is not observed between different classes (and schools), and takes the hierarchy of the data into consideration: measurements within and between students that are nested within classes that are nested within schools.

I specified a multilevel model that included all the oral and written measures (listening comprehension, receptive vocabulary, lexical richness [Guiraud Index], fluency, complexity, accuracy, grammaticality judgments). Fixed effects included main effects of AO and time as well as the interaction between AO and time. I

later added fixed effects for class size. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. Random intercepts for classes and schools were included, as were random slopes for time varying by both classes and schools, and school-specific, class-specific and task-specific slopes, using a maximal random effects structure.

The qualitative analysis of the language experience essays was conducted in two stages. The first stage involved separately reading through the essays for each student of the focus group several times, getting a general understanding of issues covered and taking note of interesting features. From the second reading on, the essays were analyzed independently by two researchers for emerging categories that were significant relative to target language development and age-related differences. 15 categories emerged as significant relative to target language development and age-related differences. Finally, after the saturation of categories, some were merged with others, resulting in eight final categories:

1. Future L2 self-states
2. Present L2 self-states
3. FL learning anxiety
4. Linguistic self-confidence
5. Attitudes towards FLs in general
6. Attitudes towards the learning situation
7. Cultural interest and media usage
8. Parental encouragement

The advantage of the conventional approach to content analysis is gaining direct information from study participants without imposing preconceived categories or theoretical perspectives. To prepare for reporting the findings, exemplars for each code and category were identified from the data.

Finally, an extensive biodata questionnaire was administered at both measurement times in order to collect biographical data and quantifiable information concerning participants' L1 and FL learning history. At the first data collection time, when the participants were under 18 years old, parents' consent was obtained to authorize the children's involvement in the research.

5. Results and analysis

5.1. Research question 1

As Table 1 in the Appendix shows, although the ECLs who took part in the longitudinal study showed stronger performance in the receptive vocabulary task as

well as with respect to oral and written lexical richness, they did not significantly outperform the LCLs in the long run with respect to receptive and productive vocabulary, and to oral and written production (content, organization, fluency, complexity, accuracy, lexical richness). The results also showed that for receptive vocabulary, grammaticality judgments, oral and written productive vocabulary (Guiraud Index), and oral and written accuracy, the Time \times AO interactions were significant in favor of LCLs, that is, the LCLs displayed faster learning rates in these areas. Not only did the LCLs make more progress within a shorter period of time in certain areas, but they were also able to catch up very quickly (i.e., within six months in secondary school) with the performance of the early starters in other areas. Thus, there was an age effect, but in favor of the late starters.

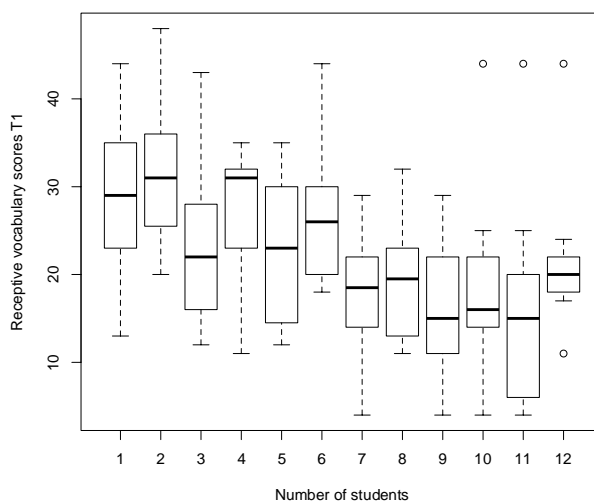


Figure 1 Variation across classes for receptive vocabulary at Time 1 (variance = 15.63, $SD = 3.60$, $p < .001$)

In addition to the fixed effects discussed above, there were also significant random class effects with estimated intra-class correlation coefficients (ICC) between 0.11 and 0.32. Class effects, therefore, explained 11%-32% of the variability in English listening comprehension, grammaticality judgments, receptive and productive vocabulary, written content, organization accuracy, fluency, complexity. Figure 1 shows the between-class differences for receptive vocabulary at Time 1.

How well a student performed in these tests was, consequently, also dependent on which class they were in—more than on the age at which they started learning English. The use of GLMs (e.g., ANOVAs, t tests) with this dataset would thus very quickly lead to incorrect estimates of treatment and other fixed effects (e.g., age effects) in the presence of the correlated errors that arise from a data hierarchy. In other words, if we fail to take the above-mentioned

variance and covariance into account statistically, this will maximize or minimize age effects, which could lead to misinformed educational policies (Goldstein, 1995; Raudenbush & Bryk, 2002). The importance of immediate context has also been observed in naturalistic studies: DeKeyser (2013), for instance, cautions that a bias in convenience samples (e.g., a bias toward the more educated or toward learners who are in contact with other speakers of the same L1) can minimize age effects in immigrant settings. Thus, to answer RQ1, classroom effects can not only impact on students' motivated behavior and, by extension, affect their FL achievements, but they also mediate age-related differences.

5.2. Research question 2

In order to clarify what exactly led to the class differences described above, I consulted the language experience essays written by the 200 subjects at both data collection times. A content analysis revealed the following factors that the students deemed conducive to FL learning at Time 1 and Time 2:

1. Group size (Time 1: ECL 65%, LCL 59%; Time 2: ECL 71%, LCL 75%): *"Our class is much too big, which doesn't honestly motivate me to contribute much to the English lesson."* (12_LLH15_M_GER)
2. Group composition (Time 1: ECL 65%, LCL 59%; Time 2: ECL 71%, LCL 75%): *"I think it's good that we only have girls in the class. We learn faster and better than other classes. My classmates spur me on."* (07_ELH21_F_GER)
3. Peer influence (Time 1: ECL 55%, LCL 50%; Time 2: ECL 33%, LCL 35%): *"A lot of my classmates thought that they [foreign languages] were sometimes boring [in primary school], so then I didn't find it fun either."* (07_ELH5_M_GER)
4. Teacher skills/personality (Time 1: ECL 79%, LCL 82%; Time 2: ECL 59%, LCL 62%): *"English was honestly not great for me from the beginning, because I didn't like our teacher so much."* (07_ELH6_F_GER)
5. Teaching method (Time 1: ECL 46%, LCL 55%; Time 2: ECL 59%, LCL 65%): *"Our English teaching was very good at primary school. We did a lot of creative stuff. And when the teaching is fun (with a lot of games too) you learn better also (I think)."* (07_ELH9_F_GER)
6. Teaching materials (Time 1: ECL 23%, LCL 25%; Time 2: ECL 12%, LCL 17%): *"If there are new modern learning methods available, they should be used! I very seldom enjoyed the French teaching . . . Besides that I find the course book 'Envol' boring and dry."* (07_LLH1_M_GER)

While factors 2-6 could not be directly measured in this study, it was possible to include class size as a fixed effect in the multilevel models. Indeed, for

all the measures, class size was a strong predictor of FL outcomes and thus partly explains why the intercepts are higher in some classes and lower in others. Figure 2 illustrates the impact of class size on receptive vocabulary at Time 2.

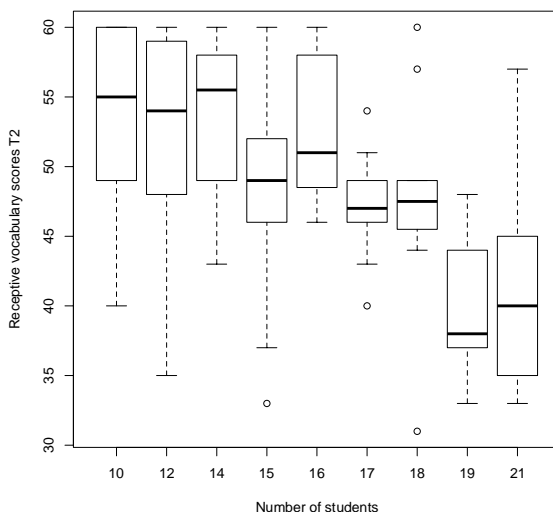


Figure 2 Effects of class size on receptive vocabulary at Time 2 ($\beta = -0.84$, $SE = 0.23$, $t = -3.66$, $p = .0006$)

Possibly one of the main reasons for this is the large impact of class size on motivation (see e.g., future L2 selves in Figure 3), which is known to mediate FL achievement (see Pfenninger & Singleton, 2016a).

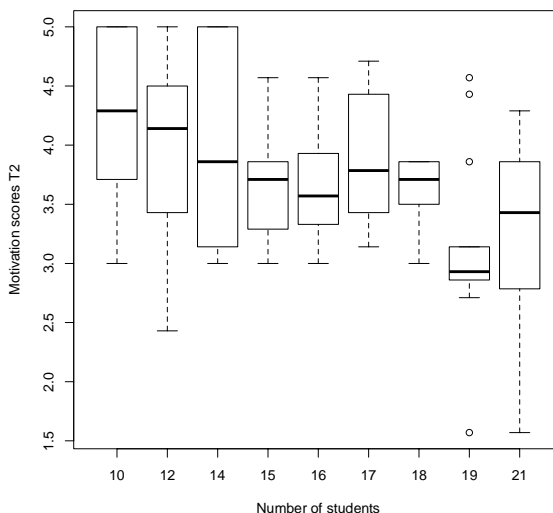


Figure 3 Effects of class size on future L2 selves at Time 2 ($\beta = -0.08$, $SE = 0.02$, $t = -3.25$, $p = .003$)

5.3. Research question 3

Multilevel analysis can also play an important role in evaluating school outcomes because it can help disentangle school effects from the effects of student characteristics (or IDs). Analyzing how much difference there was between/within schools, that is, whether there was variability in the effect of the fixed variables (AO, among others) on learners' L2 achievement, I found that there was significant variability in age effects across the five schools at Time 1 but not Time 2. Although, overall, the five schools did not vary with regard to how sensitive they were to AO across the written tasks at Time 1 (see Figure 4), the effect of AO was different for the different schools with respect to the oral measures (Figure 5) as well as various other measures (e.g., receptive vocabulary and grammaticality judgments in Figure 6 and Figure 7, respectively). Figures 4-7 thus show that some schools had weaker slopes than others for certain measures (e.g., receptive vocabulary)—meaning that age-related differences varied across schools—while for other measures (e.g., oral measures and grammaticality judgments), some schools showed age effects “in the opposite direction.”

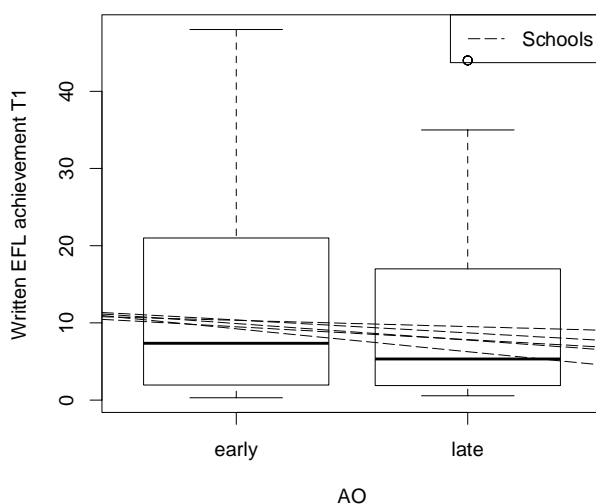


Figure 4 Random AO slopes for five schools (written EFL achievement at Time 1)

In Pfenninger and Singleton (2016), we argue that the reasons why school districts can mediate age-related differences could be the impact of schools and classes on students' motivated behavior. Furthermore, the participants came from different primary and secondary school districts and neighborhoods and hence slightly different educational backgrounds that emphasized different skills and values:

Resources available and used in FL education are dependent on schools, which might then influence learners' intrinsic interest indirectly (see e.g., Kormos & Kiddle, 2013), with the mediation of classroom factors (Muñoz, 2008). Students who are highly motivated might thus be able to make up for a later start. By the same logic, early starters who were in primary schools with less than optimal learning conditions might not be able to profit from the extended learning period, as they might have, for instance, significantly less favorable future L2 self-state. (Pfenninger & Singleton, 2016, p. 25)

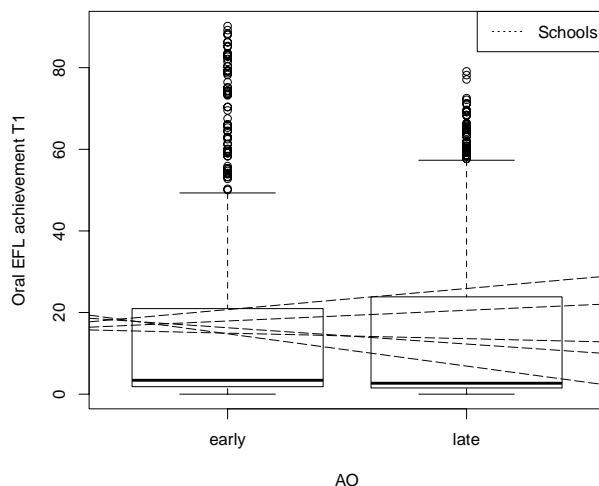


Figure 5 Random AO slopes for five schools (oral EFL achievement at Time 1)

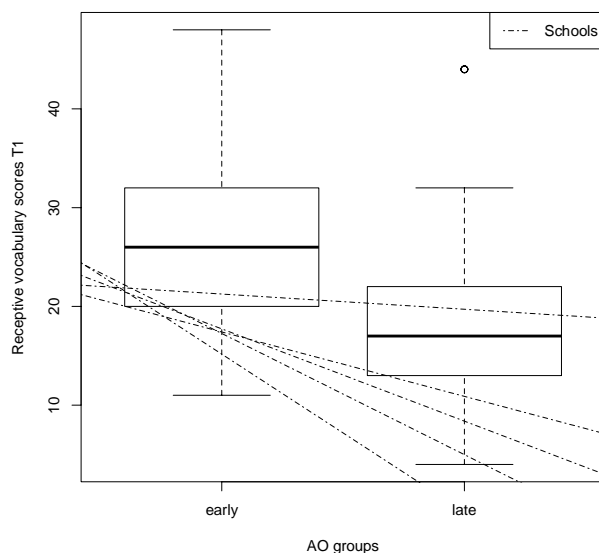


Figure 6 Random AO slopes for five schools (receptive vocabulary at Time 1)

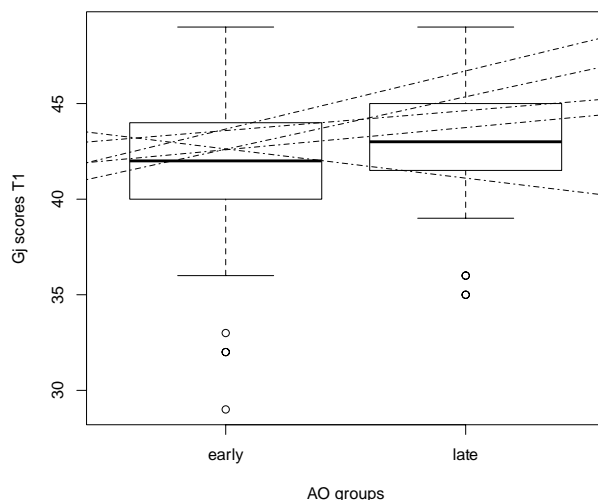


Figure 7 Random AO slopes for five schools (grammaticality judgment tasks at Time 1)

Thus, the results demonstrate how schools (in this case primary schools) can vary in their schooling outcomes, as described in the literature review above. Furthermore, the fact that school-specific slopes were no longer necessary at Time 2 shows that schools (in this case secondary schools) can also reduce inequalities between different groups over a longer period of time.

By contrast, different classes seem to be equally susceptible to age effects. Figures 8 and 9 show that some classes had a higher intercept than others, as mentioned above. Figure 8 illustrates that the earlier the students' AO is, the more the prediction for better receptive vocabulary will increase. On the other hand, late starters consistently outperformed early starters with respect to grammaticality judgments (Figure 9), arguably, because the early starters may not have developed an especially acute sense of grammatical accuracy, perhaps because of the lack of attention to this dimension in the FL instruction in primary school (see Pfenninger & Singleton, 2016). With respect to the slopes, no such significant differences can be observed. Although the slopes are not exactly parallel, the difference is relatively small, that is, learners *within* classes did not vary with regard to how sensitive they were to AO. This points to relatively strong age effects that are able to prevail despite classroom and clustering effects. However, this was a relatively small sample of five classes, and these classes had just been formed six months prior to testing, which might have had a negative impact on group cohesion (see Pfenninger & Lendl, in press). On the other hand, the findings can also be explained in terms of the strong task effects that I found, which I will discuss in the following.

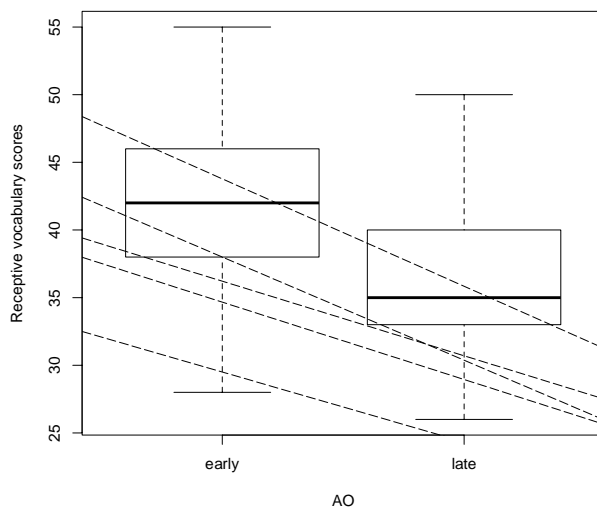


Figure 8 Random AO slopes for five mixed classes (receptive vocabulary)

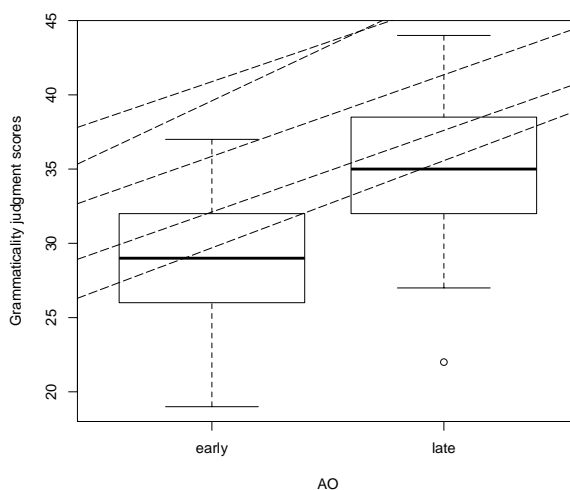


Figure 9 Random AO slopes for five mixed classes (grammaticality judgments)

In order to empirically measure and analyze whether different tasks vary with regard to how sensitive they are to the manipulation at hand (i.e., AO), I included task-specific random slopes for the fixed effect of AO so as to find out whether the effect of AO might be different for different tasks. It turned out that the effect of AO was different for different tasks at Time 1 but not at Time 2 (oral: variance = 1.43, $SD = 1.19$, $p < .001$; written: variance = 21.27, $SD = 3.5$, $p < .001$). While spoken and written fluency, complexity and accuracy as well as grammaticality judgments remain relatively unaffected by AO, receptive vocabulary (see Figure 10) and oral productive vocabulary are highly sensitive to AO. This

might reflect the greater reliance on implicit learning in children (and accordingly the implicit teaching approach in primary school) and explicit learning in older children (DeKeyser, 2012).

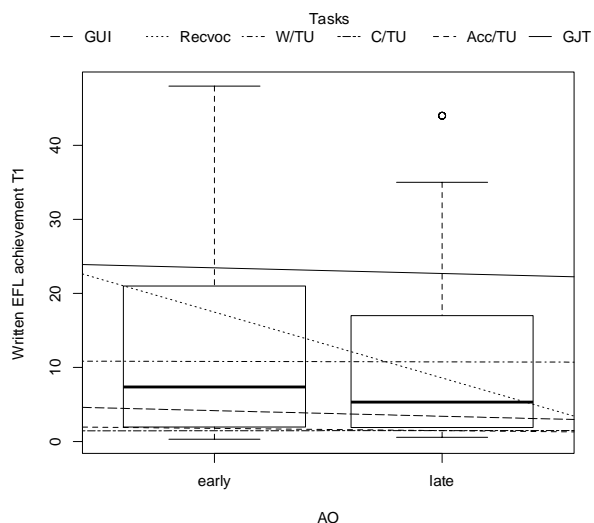


Figure 10 Random AO slopes for six written tasks (productive and receptive vocabulary, fluency, complexity, accuracy, and grammaticality judgments) at Time 1

Finally, since it is not possible to include subject-specific slopes for the fixed effect AO—which means we cannot allow the effects of AO on L2 achievement to vary across individuals in the model—I needed to employ a qualitative approach in order to find out if, for example, some ECLs profit more from an early start than others. Analyzing the language experience essays written by the focus group, that is, the 10 early high-achieving starters, the 10 early low-achieving starters, the 10 late high-achieving starters, and the 10 late low-achieving starters, revealed an interesting pattern. Although, of course, many different views and opinions emerged concerning how the students felt about the age at which they had started being exposed to English at school, there was something of a trend in that the late starters (high and low achievers alike), who had French in primary and English in secondary school, came out fairly uniformly at both data collection times (Time 1: 81%, Time 2: 91%) with critical sentiments like the following:

- (1) *I personally don't think it's good when you begin learning too early, etc. But of course I think you shouldn't start too late; I think starting English at 12 or 13 is exactly right. (07_LLH7_M_GER)*
- (2) *I think one foreign language at primary school (French) is good enough [in primary school], because we learn English anyway. I could already*

sing English songs at primary school, because I wrote them down (the words) and as a curious child I then wanted to know what it was all about. (07_LLH4_M_GER)

- (3) *The motto "the earlier the better" is not quite true. Children who don't have German as a mother tongue have to learn three languages because of this and then comes the overload. We should begin very slowly. I would leave things as they were; if I had to change something I'd put French later, from 6th class. (07_LLL6_F_GER)*

The LCLs at Time 2 on the whole remained as satisfied as they had been at Time 1 with the late English regime they had experienced, and as skeptical as they had been with regard to the wisdom of the introduction of English at primary level (see also Pfenninger, 2016). The early low achievers expressed similarly critical views at both Times (Time 1: 79%, Time 2: 86%); they mainly took issue with the slow pace in primary and the repetitions in secondary school (see Examples 4 and 5), as well as the choice of language of instruction at primary school (Example 6):

- (4) *With the help of simple games and songs in a foreign language a small vocabulary can be built up. But I remember how in early years the learning was unconcentrated and slow. At secondary level it progressed really fast. (12_ELH9_M_GER)*
- (5) *Early acquired knowledge has anyway got to be reviewed again in subsequent schooling. After five years of learning English and two years of learning French, I had to start again. (12_ELH9_M_GER)*
- (6) *At primary school our teacher even still spoke German, but here at XXX the teacher only speaks English. (07_ELb91_F_GER)*

The exception at Time 1 to the expression of dissatisfaction with what had been experienced were the early high achievers, who supported the pattern of starting English at an earlier age (Time 1: 79%).

- (7) *"The earlier the better." We should learn foreign languages early because our brain learns a foreign language faster when we're children. (07_ELH3_M_GER)*
- (8) *I think it's good that I had English as early as 2nd class because actually I didn't feel it as a burden. It was very easy too that we only learned things like "Hello, how are you" and general standard things. We learned colors, numbers and animals until finally we were able to make sentences. There were basic rules of a kind that I didn't find tremendously easy but with time you find it easier. I had a good teacher for this too. (07_ELH9_M_GER)*

At Time 2 some more nuanced, more skeptical views appeared in this group (24%), but, overall, the tenor was still in favor of an earlier start:

- (9) *Even if in individual cases early English doesn't achieve the desired success, it was still worth a try. It's of course hardly the case that children who have English instruction from second class in primary school, can speak the language fluently after four years. In my opinion, however, it's not primarily a matter of making as much progress as possible, but much more a matter of getting a feel for the language. So, for example, in relation to pronunciation and intonation.* (12_ELH6_F_GER)

Thus, the discrepancies between the groups can be ascribed to proficiency rather than AO. This hypothesis was confirmed by a majority of the participants (Time 1: ECL 52%, LCL 66%; Time 2: ECL 61%, LCL 88%) who were aware of the gap between high achievers and low achievers:

- (10) *According to my experiences, it's heavily dependent too on the person whether they benefit from the early learning of foreign languages. You have to be aware that at primary school the IQ range is very wide. So for one child French or English instruction may be a trifling thing, for another a hugely excessive demand.* (12_ELH7_F_GER)
- (11) *So actually the teaching should be suited to each child, and one group should already get foreign language teaching early and another group not yet at that time.* (12_ELH3_F_GER)

6. Conclusions and implications

It is very important to understand the true nature of age effects, not least because the age debate raises important concerns about all aspects of curriculum development and its adaptation to different ages (see DeKeyser, 2013). In this study, I have empirically measured whether AO works similarly across settings and learners or whether it is influenced by characteristics of the setting and the learner—and if so, whether there are contextual variables that can help us understand why those outcomes are different. One of the main findings was that school/class context and climate interact with student-level variables such as AO: Students under conditions of different school context and school climate demonstrate different educational attainment irrespective of AO, which has direct policy implications for policy makers, administrators, teachers, and parents. Furthermore, results of multilevel analyses indicated that macro-contextual factors (i.e., the wider school context) turn out to have a mediating effect on the

relation between AO and L2 proficiency (growth), exerting both positive and negative influences and thus suggesting malleability of AO, which is typical of ID variables. It is thus particularly important in institutional environments that age effects are considered in light of macrocultural and microcultural phenomena that can have a bearing on interpersonal relations that influence, shape, increase, or decrease variables such as motivation that interact with age.

In contrast, no such effect could be observed with lower-level data, as learners *within* classes did not vary with regard to how sensitive they are to AO, in contradistinction to other IDs such as motivation. I suggest that the origin of the significant school slopes can be found in the strong age \times context/treatment interaction documented in the literature, as well as different educational backgrounds, school curricula, materials and resources of the participants. The lack of class slopes, on the other hand, can be explained in terms of leveling effects that result from the integration of early and late starters in the same classes. The present study also showed that not only do different structures show different sensitivity to age of acquisition (see, e.g., DeKeyser, 2012) but also different tasks/skills. Arguably the focus on vocabulary in primary school is primarily responsible for this interaction effect. In the long run, however, none of the tested skills turned out to be problematic as a function of AO.

I would thus argue that the broader social environment in which learning takes place seems to be more influential than the cognitive state assumed to be a characteristic of the individual. Therefore, a simple ID model, which assumes that age is a fixed factor, is not entirely satisfactory. AO not only interacts with environmental contingencies to create a synergistic effect, but it is also influenced, mediated and mitigated by environmental influences (e.g., the impact of the learning context or compositional effects within the sample). Multilevel models are ideal for such investigations as they encourage us to shift from a myopic focus on a single factor such as the age factor to examining multiple relationships among a number of variables, including contextual variables, or, in Brown's (2011) words: "You are more likely to consider all parts of the picture at the same time, and might therefore see relationships between and among variables (all at once) that you might otherwise have missed or failed to understand" (pp. 11-12). In this view, then, such methods can be seen as an attempt to remake the connections between language learning and the social learning contexts in which these occur.

References

- Al-Thubaiti, K. (2010). *Age effects in a minimal input setting on the acquisition of English morphosyntactic and semantic properties by L1 speakers of Arabic* (Unpublished doctoral dissertation). University of Essex, UK.
- Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 161-181). Mahwah, NJ: Erlbaum.
- Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language Teaching Research, 10*, 3-32.
- Brown, J. D. (2011). Quantitative research in second language studies. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2; pp. 190-206). New York: Routledge.
- Cao, Y. (2011). Investigating situational willingness to communicate within second language classrooms from an ecological perspective. *System, 39*, 468-479.
- Chaudron, C. (2001). Progress in language classroom research: Evidence from The Modern Language Journal, 1916–2000. *Modern Language Journal, 85*, 57-76.
- Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*, 564-570.
- de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10*(1), 7-21.
- DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning, 62*(Suppl. 2), 189-200.
- DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning, 63*(Suppl. 1), 52-67.
- 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.
- Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching, 4*, 237-274.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R. (2006). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525-551). Oxford: Blackwell.

- García Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 115-135). Clevedon: Multilingual Matters.
- Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Rowley, MA: Newbury House.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Arnold.
- Granena G., & Long, M. H. (Eds.). (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29(1), 311-343. doi: 10.1177/0267658312461497
- Haenni Hoti, A., & Heinzmann, S. (2012). Foreign language reforms in Swiss primary schools – potentials and limitations. In K. Braunmüller & Ch. Gabriel (Eds.), *Multilingual individuals and multilingual societies* (pp. 189-205). Amsterdam: Benjamins.
- Harley, B. (1986). *Age in second language acquisition*. Clevedon: Multilingual Matters.
- Hyltenstam, K. (1992). Non-native features of non-native speakers: On the ultimate attainment of childhood L2 learners. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 351-368). Amsterdam: Elsevier.
- Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60-99.
- King, J. (2015). Introduction to the dynamic interplay between context and the language learner. In J. King (Ed.), *The dynamic interplay between context and the language learner* (pp. 1-10). New York: Palgrave Macmillan.
- Kinsella, C., & Singleton, D. (2014). Much more than age. *Applied Linguistics*, 35(4), 441-462.
- Kozaki Y., & Ross, S. J. (2011). Contextual dynamics in foreign language learning motivation. *Language Learning*, 61(4), 1328-1354.
- Krashen, S., Long, M., & Scarcella, R. (197). Age, rate, and eventual attainment in second language acquisition. *TESOL Quarterly*, 13(4), 573-582.
- Larsen-Freeman, D. (2015). Foreword. In J. King (Ed.), *The dynamic interplay between context and the language learner* (pp. xi-xiii). New York: Palgrave Macmillan.
- Larsen-Freeman, D., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *Modern Language Journal*, 92(2), 200-213.
- 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.
- Laufer B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, 16(1), 33-51.

- Lightbown, P. M. (2003). SLA research in the classroom/SLA research for the classroom. *Language Learning*, 28, 4-3. doi: 10.1080/09571730385200151
- Ma, X., Ma, L., & Bradley, K. D. (2008). Using multilevel modeling to investigate school effects. In A. A. O'Connell & D. B. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 59-110). Charlotte, NC: Information Age.
- 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.
- McDonald, J. L. (2008). Grammaticality judgments in children: The role of age, working memory and phonological ability. *Journal of Child Language*, 35, 247-268.
- 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.
- Montrul, S. (2008). *Incomplete acquisition in bilingualism. Re-examining the age factor*. Amsterdam: John Benjamins.
- Moyer, A. (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, C. (Ed.). (2006). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.
- Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29(4), 578-596.
- Muñoz, C. (2011). Is input more significant than starting age in foreign language acquisition? *International Review of Applied Linguistics*, 49(2), 113-133.
- Muñoz, C. (2014). Starting age and other influential factors: Insights from learner interviews. *Studies in Second Language Learning and Teaching*, 4(3), 465-484.
- Myles, F., & Mitchell, R. (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). Swindon: ESRC.
- Nikolov, M. (Ed.). (2009). *Early learning of modern foreign languages: Processes and outcomes*. Clevedon: Multilingual Matters.
- Paradis, J. (2011). Individual differences in child English second language acquisition. *Linguistic Approaches to Bilingualism*, 1(3), 213-237.
- Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, 30(2), 449-472.
- Pennycook, A. (2005). Critical applied linguistics. In L. Davis & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 783-807). Oxford: Blackwell.
- 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.

- Pfenninger, S. E. (2013). Quadrilingual advantages: Do-support in bilingual vs. multilingual learners. *International Journal of Multilingualism*, 11(2), 143-163. doi: 10.1080/14790718.2013.782032.
- Pfenninger, S. E. (2014a). The misunderstood variable: Age effects as a function of type of instruction. *Studies in Second Language Learning and Teaching*, 4(3), 415-453.
- Pfenninger, S. E. (2014b). The literacy factor in the Optimal Age Debate: A 5-year longitudinal study. *International Journal of Bilingual Education and Bilingualism*. Advance online publication. doi: 10.1080/13670050.2014.972334
- Pfenninger, S. E. (2016). All good things come in threes: Early English learning, CLIL and motivation in Switzerland. *Cahiers de l'ILSL*, 48, 119-147.
- Pfenninger, S. E., & J. Lendl (in press). Transitional woes: On the impact of L2 input continuity from primary to secondary school. *Studies in Second Language Learning and Teaching*.
- Pfenninger, S. E., & Singleton, D. (2016). Affect trumps age: A person-in-context relational view of age and motivation in SLA. *Second Language Research*. Advance online publication. doi: 10.1177/0267658315624476
- Pfenninger, S. E., & Singleton, D. (2017). *Beyond age effects in instructional L2 learning: Revisiting the age factor*. Bristol: Multilingual Matters.
- Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, 35, 655-687.
- Plonsky, L. (2014). Study quality in quantitative L2 research (1990-2010): A methodological synthesis and call for reform. *Modern Language Journal*, 85, 450-470.
- Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, 61, 325-366.
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Robinson, P. (2002). *Individual differences and instructed language learning*. Amsterdam: John Benjamins.
- 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.
- Singleton, D., & Ryan, L. (2004). *Language acquisition: The age factor* (2nd ed.). Clevedon: Multilingual Matters.
- Snow, C., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, 49, 1114-1128.
- Stevens, G. (2006). The age-length-onset problems in research on second language acquisition among immigrants. *Language Learning*, 56(4), 671-692.

- Sze, P. (1994). The optimal starting age for second language learning: The case of Hong Kong. *CUHK Journal of Primary Education*, 4(2), 49-54.
- Unsworth, S., de Bot, K., Persson, L., & Prins, T. (2012). Foreign languages in primary school project. *Proceedings of the Foreign Languages in Primary Schools Projects: Presentation results FLiPP-research*.
- Ushioda, E. (2008). Motivation and good language learners. In C. Griffiths (Ed.), *Lessons from good language learners* (pp. 19-34). Cambridge: Cambridge University Press.
- van Geert, P., & Steenbeek, H. (2008). A complexity and dynamic systems approach to development assessment, modeling and research. In A. M. Battro, K. W. Fischer, & P. Léna (Eds.), *The educated brain* (pp. 71-94). Cambridge: Cambridge University Press.
- Vanhove, J. (2013). The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PloS One*, 8(7). doi: 10.1371/journal.pone.0069172
- van Lier, L. (1988). *The classroom and the language learner*. New York: Longman.
- Wen, W. P., & Clément, R. (2003). A Chinese conceptualization of willingness to communicate in ESL. *Language, Culture and Curriculum*, 16(1), 18-38.
- Zafar, S., & Meenakshi, F. (2012). Individual learner differences and second language acquisition: A review. *Journal of Language Teaching and Research*, 3(4), 639-646.

APPENDIX

Table 1 Evaluation of written production and response

	Time 1					Time 2				
	ECL ₁	LCL ₁	β	t	p	ECL ₂	LCL ₂	β	t	p
Listening	n.a.	n.a.	n.a.	n.a.	n.a.	12.61 (3.17)	12.08 (3.49)	-0.45 ± 0.47	-0.95	.332
Productive vocabulary	n.a.	n.a.	n.a.	n.a.	n.a.	25.30 (7.35)	26.18 (7.88)	0.94 ± 2.83	0.33	.716
Receptive vocabulary	26.36 (8.59)	17.47 (8.05)	-8.90 ± 1.73	-5.13	<.001*	50.08 (7.14)	49.40 (7.54)	-0.60 ± 3.28	-0.18	.841
Written content	19.14 (2.61)	19.05 (2.19)	-0.10 ± 0.34	-0.29	.791	27.27 (1.91)	27.10 (2.01)	-0.20 ± 0.37	-0.54	.567
Written organization	10.61 (2.16)	10.42 (2.05)	-0.10 ± 0.30	-0.35	.709	16.67 (2.96)	16.90 (2.45)	0.28 ± 0.49	0.58	.560
Guiraud index (oral)	4.05 (1.90)	3.21 (1.29)	-0.97 ± 0.96	-2.01	.002*	5.55 (1.43)	5.63 (1.27)	-0.00 ± 0.83	-0.00	.997
Guiraud index (written)	4.92 (1.30)	4.17 (0.78)	-0.76 ± 0.19	-4.09	<.001*	7.57 (0.80)	7.73 (0.77)	0.16 ± 0.15	1.04	.268
Fluency (oral)	60.95 (16.55)	58.00 (8.28)	-5.03 ± 7.08	-0.63	.494	124.80 (12.78)	122.63 (12.92)	-2.23 ± 7.41	-0.30	.742
Fluency (written)	10.87 (3.64)	10.78 (3.22)	-0.09 ± 0.49	-0.19	.846	14.91 (2.97)	14.21 (4.17)	-0.74 ± 0.57	-1.31	0.18
Complexity (oral)	1.32 (0.62)	1.34 (0.41)	-0.05 ± 0.31	-0.17	.862	1.57 (0.50)	1.61 (0.50)	0.00 ± 0.28	0.00	.996
Complexity (written)	1.43 (0.39)	1.45 (0.31)	-0.00 ± 0.05	-0.03	.996	1.69 (0.61)	1.71 (0.44)	-0.01 ± 0.09	-0.14	.900
Accuracy (oral)	3.46 (1.67)	2.79 (1.72)	-0.67 ± 0.24	-2.78	.008*	1.20 (1.25)	1.30 (1.40)	0.04 ± 0.19	0.25	.763
Accuracy (written)	2.07 (0.63)	1.77 (0.58)	-0.33 ± 0.08	-4.03	<.001*	0.60 (0.44)	0.62 (0.56)	0.02 ± 0.08	0.30	.745
Grammaticality judgments	24.20 (3.78)	23.45 (3.41)	-0.79 ± 0.66	-1.19	.203	41.93 (3.31)	42.97 (2.75)	0.96 ± 0.76	1.27	.180

Note. *Statistically significant at $\alpha < .05$; bold type = significantly higher scores.