Despite contrary research findings, many laypeople still claim that starting second language (L2) instruction early yields linguistic advantages. This assertion is again undermined by a 5-year longitudinal study conducted in Switzerland testing the English language skills of 636 secondary-school students who had all learned Standard German and French at primary school, but only half of whom had learned English from age 8, the remainder having started English 5 years later. The results suggest that age-related attainment effects are overshadowed by other effects, yielding diverse outcomes according to individual differences and contextual effects mediating L2 outcomes. An earlier age of learning proved beneficial only for children raised as biliterate simultaneous bilinguals receiving substantial parental support, as opposed to monolinguals and nonbiliterate bilinguals (simultaneous or sequential). These issues require studies that explore what underlies age effects in L2 learning and investigate how learning contexts shape processes of L2 development.

**Keywords** age factor; multilingualism; bilingual advantage; young learners; early foreign language learning
Introduction

Research has for more than half a century consistently failed to find any long-term advantage to beginning second language (L2) learning at primary as compared to secondary school (e.g., see Muñoz & Singleton, 2011; Pfenninger & Singleton, 2017). Furthermore, age-related classroom research has been prey to a monolingual bias in that age researchers have focused on the comparison of monolingual early and late starters (e.g., see García Mayo & García Lecumberri, 2003; Muñoz, 2006)—despite the fact that over 50% of the population of the European Union, for example, report being functionally bilingual (European Commission, 2011). Such figures often imply the presence of three or more languages in primary foreign language (FL) classes. According to the European Agency for Development in Special Needs Education (2003), primary-school teachers in Europe see their classes becoming more linguistically heterogeneous, and multilingual schools are growing increasingly common. Moreover, pupils bring into their classes home languages that often differ from the school languages (Muñoz & Singleton, in press).

The skill development profile derived from experience with monolingual students may well differ from that of children establishing basic cognitive competencies in childhood through the mediation of two languages (see Bialystok & Feng, 2011). Bilingualism has been linked to the enhancement of domain-general executive control abilities and of the additional language learning process (Bialystok, Abutalebi, Bak, Burke, & Kroll, 2016; Bialystok, Craik, & Luk, 2008). The general validity of the claim that bilinguals are better language learners, however, is difficult to establish because of variation among bilinguals and the social dimension of bilingualism.

Another issue vis-à-vis primary-school language learning is how much students bring to this from their previous experience. Many of us would agree that “the child’s existing knowledge and experience form the basis for learning” (Government of Ireland, 1999, p. 8) and that “parents are the child’s primary educators” (p. 24), especially because this postulate has plentiful research backing (see Goodall et al., 2011). We might then see family circumstances as influencing different learning outcomes.

The present study focuses on (a) learners of English of different starting ages raised in Germanophone Switzerland with German (Swiss and Standard) and who had then learned English and (b) learners of English of different starting ages who had either been born in Switzerland and brought up from birth with German plus another language or else had arrived in Switzerland from another language background at age 5 or 6 and had acquired German in addition to their L1 in mid-childhood. We call the former “monolingual” learners of
English, ignoring for present purposes the fact that Swiss German is very different from Standard German as well as the fact that the children had some school contact with French. The latter group we label “bilingual,” categorizing them into three subsets according to their experience, once again simplifying things terminologically by glossing over the Swiss/Standard German distinction and the (minimal) French connection. The study investigates the extent to which being (in the present sense) bilingual or biliterate improved attainment in English and also the role of contextual factors and how these predictors interact with starting age effects.

**Effects of Age of First FL Instruction: Some Background**

**Age As a Slippery Individual Difference Variable**

Age is usually viewed as an individual difference—along with abilities, propensities, learner cognitions/beliefs about L2 learning, and learner actions/strategies. Ellis (2006), however, disagrees with this view: “‘[A]ge’ itself does not belong to any of the four categories; rather, it potentially affects learners’ abilities, propensities, cognitions, and actions (as do other factors such as previous learning experiences and the learning situation)” (p. 529). He argued that “the role played by age in L2 acquisition warrants an entirely separate treatment” (p. 530) from individual differences.

For some time, the age factor has been recognized as a “macrovariable” (Flege, Yeni-Komshian, & Liu, 1999), and age researchers have been calling for dimensions other than maturation to be taken into consideration. Advocates of the Critical Period Hypothesis (CPH) (like Montrul—e.g., 2008, 2012) and CPH skeptics (like Moyer—e.g., 2013, 2014a, 2014b) are one on this point. Montrul (2008) declared that the age notion “cannot be isolated from other co-occurring factors” (p. 50). According to Moyer (2013), interrelated variables are at play, “having to do with learner orientation and experience” (p. 1). The general view is that researchers need to attend to age-related social, psychological, and contextual factors—such as positive attitudes, risk-taking behavior, openness to the new culture, greater commitment of time and/or energy, identity construction, better support system, friendships with speakers of the target language, urge or need to fit in, exposure to rich and varied input in school, and so on—as well as to strictly maturational factors.

The relationship between age and additional language acquisition is complex. DeKeyser (2013) claims that a bias in sampling can minimize age effects in immigrant settings. If convenience samples favor more educated participants or are taken from communities that comprise many native speakers of a certain language, this can influence the age-related outcome. Similarly, Pfenninger
(2017) demonstrated for the classroom setting how the characteristics of the groups investigated have implications not only for theoretical discussions of the age factor but also for methodology. Compositional effects within the sample (students nested within classes nested within schools nested within school districts, etc., or children coming from different families, neighborhoods, etc.) impact and mediate age effects, which tend to disappear when the clustered data are taken into account. This is because as soon as individuals become part of a group, they accommodate to the normative environment of that group. This normalizing influence may augment or undermine learners’ individual motivations and hence proficiency level (see Pfenninger & Singleton, 2016). Age of onset (AO) thus not only interacts with environmental contingencies to create a synergistic effect, but it is also influenced, mediated, and mitigated by such environmental influences. Although cohort effects have been well documented, relevant observations have been typically neglected in analyses; that is, age researchers have tended to decontextualize their findings. Likewise, the home context has usually been neglected in age studies.

**Effects of Family Circumstance in Education**

The literature concerning parental influence on general academic motivation and performance among children is rich and positive. One important research survey (Goodall et al., 2011) concluded baldly that “[p]arental engagement has a large and positive impact on children’s learning” (p. 3). Wigfield, Eccles, Schiefele, Roeser, and Kean (2006) identified four influential parental factors: (a) parental, familial, and neighborhood characteristics; (b) parents’ general beliefs and behaviors; (c) parents’ child-specific beliefs; and (d) parent-specific behaviors. Parental support has been interpreted as influencing children’s school performance (Fan & Chen, 2001) at both primary and secondary levels (Feinstein & Symons, 1999), having the capacity positively to affect academic achievement, cognitive competence, problem-solving skills, school enjoyment, school attendance, and behavior (Melhuish, Sylva, Sammons, Siraj-Blatchford, & Taggart, 2001). A home environment where the child is given exposure to academically oriented vocabulary and books is found to positively influence the child’s school achievement, a phenomenon that includes nonnative learners (e.g., see Goldenberg, Rueda, & August, 2008).

With regard to family support effects on the quality of the language learning process, the findings are in line with those of general educational research. For example, Csiszér and Kormos (2009) found that one of the factors that appeared to affect Hungarian students’ L2 learning attitude and motivation was parental encouragement. Morris, Lafontaine, Pichette, and de Serres (2013) also
found in relation to L2 learning among South Korean students that parental encouragement correlated with motivation to learn and parental disinterest correlated negatively with motivation. Castillo and Camelo Gámez’s (2013) study suggested that giving parents some instruction in regard to assisting their children with L2 learning produce positive results, yielding good effects on the children’s L2 learning and also contributing to their literacy development.

Similarly, with regard to parental input in the rearing of multilinguals, according to McCabe et al. (2013), parents of multilingual children who talk at length to their children, ask them numerous questions, treat new vocabulary in meaningful context, and bookshare in various ways have a substantive positive impact on their children’s linguistic development. Also, in regard to home multilingual development, Duursma et al. (2007) found a sibling effect: Language preference for interaction with siblings had an even larger effect on respective language proficiency than parental language preference. Family circumstances may also influence bilingualism at school: Goriot, Denessen, Bakker, and Droop (2016) found that the more positive children’s perceptions of tolerance (i.e., teachers’ appreciation of the home language) were (perhaps deriving from their positive family experience), the better they performed on nonverbal working memory tasks.

**Age and the Bilingual Advantage**

As noted earlier, bilingualism has been related to the enhancement of central executive control functions. As people who are bilingual constantly need to juggle both languages (see Green, 1998), inhibiting one when speaking the other, they allegedly have an enhanced inhibitory control mechanism, which allows them to regulate and control their attention in other (nonlinguistic) cognitive tasks (see Adesope, Lavin, Thompson, & Ungerleider, 2010; Bialystok et al., 2016). Bilingualism has also been hypothesized to result in more efficient language learning, in terms of the attainment of both general language proficiency (Cenoz & Valencia, 1994; Swain, Lapkin, Rowen, & Hart, 1990) and of literacy skills (e.g., Kovelman, Baker, & Petitto, 2008). However, a whole body of evidence questioning the notion of a general bilingual advantage has emerged recently, relating to:

- the hybridity of experiences of bilinguals in these studies, which may be associated with a subject selection bias (De Angelis, 2015);
- the social dimensions of bilingualism, that is, the influence of and changes in lifestyle, L2 learning motivation, overall well-being, general communicative skills, the status/prestige of the languages in question, teachers’ cultural responsiveness, and so forth (e.g., Agirdag, 2010; Goriot et al., 2016);
• a priori cognitive ability (Keijzer & Schmid, 2016);
• methodological inconsistencies, for example, the differential effects of bilingualism on verbal and nonverbal task performance (Duñabeitia & Carreiras, 2015; Paap, Johnson, & Sawi, 2015; Vaughn, Greene, Ramos Nuñez, & Hernandez, 2015); and
• an alleged publication bias favoring positive outcomes over null effects and possibly leading to a false representation of the true situation (see Bialystok, Kroll, Green, MacWhinney, & Craik, 2015; de Bot, 2017; de Bruin, Treccani, & Della Sala, 2015).

De Bot (2017) suggested that “rather than one all-purpose bilingual advantage, there may be different advantages—but also disadvantages—for specific groups (early/late bilinguals, migrants, high/low education) and specific cognitive functions (working memory, attention, task switching)” (p. 25).

Some scholars have suggested that the key to the cognitive advantage reflected in more efficient L3 acquisition is biliteracy, rather than exclusively oral bilingualism (Rauch, Naumann, & Jude, 2012; Sanz, 2008; Swain et al., 1990). Bilinguals illiterate in their L1 do not perform significantly better than their monolingual counterparts. In Sanz’s (2008) study of 120 Catalan-Spanish bilingual adolescents learning L3 English, for instance, balance and biliteracy were predictive of performance on various measures of L3 grammatical proficiency. Sanz (2008) concluded that a “degree of bilingual literacy is key to success in L3 acquisition” (p. 235) in that biliterate bilinguals outperform nonbiliterate bilinguals on grammar tests (although not on vocabulary tests). From a socio-affective perspective, various studies have shown that a strong basis in the L1 promotes school achievement in the L2 and that it is important in this context to ensure that children do not become alienated from their families and communities (Castro, Páez, Dickinson, & Frede, 2011; Murphy & Evangelou, 2016).

Others suggest that it takes a lifetime of communicating in two or more languages from an early age for the bilingual advantage to unfold (see, e.g., Bialystok, Craik, Binns, Ossher, & Freedman, 2014; Bialystok, Craik, & Freedman, 2007; Perquin et al., 2013). Some studies have found that advantages also emerge for bilingual populations who are less experienced in one of their languages (e.g., Bialystok & Martin, 2004). Thus, Sanz (2008) observed that simultaneous bilinguals did not show an advantage over those who had learned the L2 at age 7, after becoming literate in their L1. On the other hand, some researchers have demonstrated negative correlations between onset age of active bilingualism and English FL proficiency. Luk, de Sa, and Bialystok (2011), for instance, studied the relationship between onset age
of bilingualism and cognitive control, comparing early bilinguals (actively bilingual before age 10), late bilinguals (actively bilingual after age 10), and monolinguals. Early bilinguals and monolinguals demonstrated similar levels of English proficiency, and both groups were more proficient in English than the late bilinguals (see also Soveri, Rodriguez-Fornells, & Lain, 2011). Interestingly, Stafford, Sanz, and Wood Bowden (2010) found a tendency for late bilinguals to develop higher proficiency than early bilinguals in adult FL learning. The authors detected “a slight advantage” (p. 179) for late bilinguals over early bilinguals with respect to ability to retain what the participants learned about the new and complex noun case morphology cue. They ascribed this finding to the highly structured and explicit instructional treatment that might have been more in tune with late bilinguals’ learning strategies.

Given that different types of bilingual experience may affect the development of executive control in different ways and to different extents (Yow & Li, 2015), it is worth making a finer-grained distinction between different types of bilinguals who begin FL instruction earlier and later in life (or school), respectively. The current state of the art also calls for investigations of whether early bilinguals also enjoy more successful outcomes than monolinguals in a school context and whether these bilingual advantages also emerge for late (sequential) bilinguals.

The Present Study

To generate richer and better-supported insights into the age factor in second language acquisition, we conducted a longitudinal study “Beyond Age Effects” (2009–2016), addressing (inter alia) the following research questions (see also Pfenninger, & Singleton, 2017):

1. How does age of FL learning onset affect different learner populations in the short and the long run?
2. What is the relationship between direct and indirect family support for language learning and the children’s bilingualism/biliteracy and learning of English?

We also demonstrate methods that we see as ideal for studying age effects, including statistical models that can maximize the generalizability of the findings without overly sacrificing variability.

Participants and Procedure

The participants were 636 learners of English as a foreign language (EFL) living in the Zurich area of Switzerland, 325 of whom were early classroom
learners (ECLs; AO 8, Year 2) and 311 late classroom learners (LCLs; AO 13, Year 7). Thus, at the first measurement, the ECLs had already received 5 years of EFL instruction, while the LCLs had only just begun learning English 6 months before. We chose four samples that represent four important sectors of the school-age population in Switzerland:

- **Group MONO**: 200 Swiss monolinguals, born in Switzerland (speakers of Swiss German, literate in the community language of literacy, Standard German)
  - 100 ECLs (AO 8)
  - 100 LCLs (AO 13)

- **Group SIMBI I**: 144 simultaneous bilinguals, born in Switzerland (proficient in Swiss German; biliterate—i.e., literate in the community language of literacy, Standard German, and also literate in their other language)
  - 73 ECLs (AO 8)
  - 71 LCLs (AO 13)

- **Group SIMBI II**: 107 simultaneous bilinguals, born in Switzerland (proficient in Swiss German, literate in the community language of literacy, Standard German, illiterate in their other language)
  - 57 ECLs (AO 8)
  - 50 LCLs (AO 13)

- **Group SEQBI**: 185 sequential bilinguals, age of arrival in Switzerland 5–6 (illiterate in their L1; proficient in Swiss German, literate in the community language of literacy, Standard German)
  - 95 ECLs (AO 8)
  - 90 LCLs (AO 13)

All participants were within the age range 13–14 years (mean age 13.4) at the first data collection time in 2009 and in the range 18–19 years (mean age 18.6) at the second measurement in 2014; they were learning English as an L3 and L2 of literacy (in the case of the monolinguals) or as an L4 and L2 and L3 of literacy (in the case of the bilinguals). In each group, roughly half the participants were ECLs, while the other half were LCLs.

As for the teaching approach and classroom practices in primary versus secondary school, implicit, communicative, and holistic approaches are the norm in early foreign language instruction in the Swiss primary-school classroom;
in other words, young children under the age of 10 or so are for the most part typically taught in a playful way via songs, games, and so forth, with a focus on oral use of the L2. Secondary school, by contrast, is characterized by more explicit focus on form. All the children had received initial literacy instruction in Standard German. For the sequential bilinguals, therefore, the language of schooling was not the same as the home language. They acquired German language literacy in school in spite of having a weak command of spoken German. Standard German was not spoken at home or in the community either.

Because very early exposure to the second, majority language seems to affect L1 skills (e.g., Cobo-Lewis, Pearson, Eilers, & Umbel, 2002), it was decided to include only individuals in the SEQBI group who had moved to Switzerland between the ages of 5 and 6. During their first 5 years they were exposed only to their L1 and acquired some (modest) preliteracy skills that were not maintained after they began training in the majority language of literacy (Standard German). Answers to an extensive biodata questionnaire completed by all participants confirmed that the bilingual groups did not start school behind their monolingual peers, and their families were not disproportionately of low income compared to the monolingual families.

As in, for example, Cenoz and Valencia (1994) and Sanz (2000), language typology is not a variable scrutinized in this study. We ensured, however, that the participants in the two bilingual groups shared the same array of L1 backgrounds: Spanish, Portuguese, Croatian, Serbian, Albanian, Arabic, or Italian. (None of these languages features in the school curriculum.) In other words, the same number and types of language pairs (e.g., German–Albanian, German–Serbian) feature in each bilingual group. Nevertheless, it is clear that the relationships among the languages involved might also play some role in accounting for differences in the results of this kind of research.

The study does not operationalize level of bilingualism either. Through written questionnaires, parents confirmed that children were only proficient in German and one of the seven languages listed above and used German and their home language regularly. While we included only children considered functionally bilingual, they might have had slightly different levels of proficiency in their different languages.

Finally, information about parents’ occupation and education level was also obtained via this questionnaire. While parents’ English proficiency level was included as a fixed factor in the statistical models (see below), socioeconomic status (SES) was not specified as a fixed factor for the following reasons: SES discrepancies are not as pronounced in Switzerland as in other European countries, although discrepancies exist between the education of foreign and
native children (see Organisation for Economic Co-operation and Development, 2007); also, answers to an extensive biodata questionnaire completed by all participants confirmed that the bilingual groups did not start school behind their monolingual peers, and their families were not disproportionally low income compared to the monolingual families.

**Measures**

*Sociolinguistic Context Measures*

Children were tested within regular classroom hours over 2 weeks at each data-collection time. A parental questionnaire was also developed consisting of 13 items. This focused on environmental factors previously found to influence children’s achievement and motivation (see above):

- **Books/e-books:** ordinal measure ranging from 1 to 3
  - between 0 and 10 books in the home
  - between 11 and 50 books in the home
  - between 51 and 100 books in the home

- **Family support**
  - Indirect behaviors:
    - Parents’ English proficiency: self-assessment of listening, speaking, reading, and writing on a 5-point Likert scale (the average scores for the four skills were used)
    - Presence of older siblings with positive attitudes to EFL: binary, 0 = no, 1 = yes
  - Direct behaviors:
    - Parents’ direct assistance in helping their children learn/study English, such as helping them with their homework (4 items in a 5-point ordinal measure: 5 = daily, 4 = three times a week, 3 = once a week, 2 = once a month, 1 = never).

- **Attitudes and beliefs of parents**
  - Parents’ attitudes toward EFL: perceptions in relation to the role of English (2 items, a 5-point Likert scale)
  - Parents’ opinion of the child’s English education (2 items, a 5-point Likert scale)
  - Parents’ beliefs and expectations regarding their child’s abilities/success in acquiring English (2 items, a 5-point Likert scale)
EFL Proficiency Measures

The following tasks and instruments were administered to measure linguistic skills (all pilot-tested and used in Pfenninger & Singleton, 2017):

- Listening comprehension: 2 standardized listening comprehension tasks (Common European Framework of Reference for Languages [CEFR] level B2)
- Written lexical richness, syntactic complexity, fluency, and accuracy: English argumentative essay on the pros and cons of (reality TV) talent shows
- Oral lexical richness, syntactic complexity, fluency, and accuracy: Oral tasks (retelling task, spot-the-difference task)
- Grammaticality judgments: Grammaticality judgment task including 49 items and 15 distractors (reliability coefficient KR-20 .90 for grammatical items and .95 for ungrammatical items)

Because the achievement levels for the end of compulsory school is CEFR B1 level in Switzerland (Bildungsdirektion des Kt. Zürich, 2010), the tests aim to measure B1–B2 levels.

Furthermore, as it has been suggested that there may be different windows for different language domains (e.g., Granena & Long, 2013) and that, consequently, effects of AO may be different for different tasks (see Pfenninger, 2017), we made use of multiple elicitation methods. Two of the tasks (listening comprehension task, productive vocabulary test) could only be administered once because the pilot testing showed that those measures were appropriate for the students at Time 2 but would have yielded floor effects at Time 1.

Oral and written competence was measured in terms of fluency, lexical and syntactic complexity, and accuracy. Following Wolfe-Quintero, Inagaki, and Kim (1998), fluency in English and Standard German was examined in terms of words per T-unit (W/TU), defined as one main clause and all dependent modifying clauses (Ellis & Barkhuizen, 2005). Oral and written syntactic complexity was examined in English and Standard German using the clauses per T-unit (CL/TU) complexity ratio. Lexical complexity in the oral and written data was examined using Guiraud’s Index of Lexical Richness (GUI): word types divided by the square root of the word tokens. Accuracy (ERR/TU)
was examined by counting (a) misspellings (excluding mechanical errors such as punctuation errors) and (b) the number of morphosyntactic errors per T-unit.

All tasks, described in detail below, were piloted in two classes of intermediate learners (n = 54, mean age 15.9 years) in June and August 2008 and—if necessary—revised in the light of the piloting results to elicit the intended kind of responses and to obtain the appropriate difficulty level for the students concerned and ensure sufficient validity and reliability coefficients (see Pfenninger & Singleton, 2017).

**Data Analysis**
Analyses were conducted using mixed-effects models with hierarchical random effects for classes and schools, and crossed random effects for subjects and items, respectively, using the lme4 package (Version 1.1–7) in R (Version 3.1.0; R Development Core Team). While continuous fixed effects were centered, categorical fixed effects were recoded to use contrast coding (e.g., Linck & Cunnings, 2015). The final models included contrast coded fixed effects for AO (−.5 = early, .5 = late), bilingualism (−.5 = nonbilingual, .5 = bilingual), biliteracy (−.5 = nonbiliterate, .5 = biliterate), type of bilingualism (−.5 = simultaneous, .5 = sequential), as well as time (for the growth analyses), number of books/e-books, degrees of in/direct parental involvement, and frequency of reading. Random effects were fitted using a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013; but for recent challenges to this approach, see Bates, Kliegl, Vasishth, & Baayen, 2015, and Baayen, Vasishth, Bates, & Kliegl, 2015). We included 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.

Because the same items in the listening comprehension task, receptive vocabulary test, productive vocabulary test, and grammaticality judgment task were presented to both groups of participants (ECLs and LCLs), we included a random slope (AO|item) in each of these models, which allow the AO effect to vary by item. Furthermore, because time varied within subjects, we added a random slope of time, varying by student (random slopes for classes and schools were unnecessary).

In those cases where the dependent variables were not continuous but categorical/binary—for example, grammaticality judgments (correct/incorrect)—we used the mixed-effects implementation of logistic regression or mixed logit models. In the case of count data (number of errors in the accuracy measure and counts of words and clauses in the fluency
and complexity measures, respectively), which were not normally distributed either, a generalized mixed-effects model (glmer) with the Poisson distribution was used. We used visual inspection of residual plots to find if there were any obvious deviations from homoskedasticity or normality. Models were fitted using a maximum likelihood technique. \( p \) values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. Besides reporting means and standard deviations, we also indicate confidence intervals (CIs) at a probability level of 95% to contain the true population mean and provide a reliable way to indicate whether the difference between a pair of mean scores is statistically significant and, importantly, whether the difference is stable.

**Results**

The first research question asked how age of FL acquisition onset affects different learner populations in the short and the long run. The performance of the ECLs and LCLs of each language group (MONO, SIMBI I, SIMBI II, and SEQBI) across a range of EFL skills at each data collection time (T1 at the beginning of secondary school and T2 5 years later at the end of mandatory school time) is summarized in Appendix S1 in the Supporting Information online.

Mixed models were specified for each of the 12 FL measures to examine the predictive power of AO, bilingualism, biliteracy, and contextual factors (number of books in the household, direct and indirect family support, and parental attitudes). In all four groups, an earlier AO was a significant predictor of 60% of the tested FL skills at the beginning of secondary school (see Table 1). Specifically, in each group, the early starters outperformed the late starters in receptive vocabulary, written lexical richness, written fluency, oral lexical richness, oral accuracy, and written grammaticality judgments. When comparing the scores of the ECLs to the CIs for the LCLs, the ECLs’ performance fell outside of the CI of the LCLs for all of these measures. However, parity between early starters and late starters had already been reached after 6 months of secondary-school EFL instruction in terms of written and oral complexity, written accuracy, and oral fluency.

At the end of mandatory school time these AO effects had been effectively washed out, that is, only written lexical richness still reached significance (see also Table 1). The difference between the ECLs’ scores and the LCLs’ scores can no longer be deemed significant considering the CIs either. From this it follows that the LCLs made more progress than the ECLs over the period of 5 years in secondary school. Table 2 shows that for receptive vocabulary, oral
Table 1: Multilevel regression analyses for the dependent variables at Time 1 (T1) and Time 2 (T2) (fixed-effect estimates for age of onset [AO])

<table>
<thead>
<tr>
<th>Measure</th>
<th>Estimate ± SE</th>
<th>t</th>
<th>Main effect p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed effect: AO at T1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Receptive vocabulary</td>
<td>−1.73 ± 1.83</td>
<td>−5.95</td>
<td>&lt; .0001**</td>
</tr>
<tr>
<td>Written lexical richness</td>
<td>−0.62 ± 0.15</td>
<td>−4.22</td>
<td>.0001**</td>
</tr>
<tr>
<td>Written fluency</td>
<td>−0.54 ± 0.27</td>
<td>−2.01</td>
<td>.041*</td>
</tr>
<tr>
<td>Written complexity</td>
<td>−0.00 ± 0.03</td>
<td>−0.12</td>
<td>.913</td>
</tr>
<tr>
<td>Written accuracy</td>
<td>0.01 ± 0.10</td>
<td>0.13</td>
<td>.914</td>
</tr>
<tr>
<td>Oral lexical richness</td>
<td>−0.54 ± 0.33</td>
<td>−2.07</td>
<td>.049*</td>
</tr>
<tr>
<td>Oral fluency</td>
<td>−2.82 ± 2.95</td>
<td>−0.96</td>
<td>.322</td>
</tr>
<tr>
<td>Oral complexity</td>
<td>−0.04 ± 0.09</td>
<td>−0.51</td>
<td>.592</td>
</tr>
<tr>
<td>Oral accuracy</td>
<td>−0.35 ± 0.15</td>
<td>−2.36</td>
<td>.018*</td>
</tr>
<tr>
<td>Grammaticality judgment task</td>
<td>−1.12 ± 0.41</td>
<td>−2.73</td>
<td>.007**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure</th>
<th>Estimate ± SE</th>
<th>t</th>
<th>Main effect p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed effect: AO at T1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening</td>
<td>−0.33 ± 0.37</td>
<td>−0.90</td>
<td>.317</td>
</tr>
<tr>
<td>Productive vocabulary</td>
<td>−0.85 ± 1.65</td>
<td>−0.51</td>
<td>.372</td>
</tr>
<tr>
<td>Receptive vocabulary</td>
<td>−0.86 ± 1.14</td>
<td>−0.76</td>
<td>.415</td>
</tr>
<tr>
<td>Written lexical richness</td>
<td>−0.50 ± 1.17</td>
<td>−2.95</td>
<td>.012*</td>
</tr>
<tr>
<td>Written fluency</td>
<td>−1.28 ± 0.73</td>
<td>−1.74</td>
<td>.193</td>
</tr>
<tr>
<td>Written complexity</td>
<td>−0.00 ± 0.05</td>
<td>−0.17</td>
<td>.851</td>
</tr>
<tr>
<td>Written accuracy</td>
<td>−0.01 ± 0.05</td>
<td>−0.27</td>
<td>.840</td>
</tr>
<tr>
<td>Oral lexical richness</td>
<td>0.03 ± 0.31</td>
<td>0.08</td>
<td>.927</td>
</tr>
<tr>
<td>Oral fluency</td>
<td>−2.63 ± 3.20</td>
<td>−0.82</td>
<td>.387</td>
</tr>
<tr>
<td>Oral complexity</td>
<td>−0.19 ± 0.11</td>
<td>−1.71</td>
<td>.132</td>
</tr>
<tr>
<td>Oral accuracy</td>
<td>0.05 ± 0.16</td>
<td>0.31</td>
<td>.750</td>
</tr>
<tr>
<td>Grammaticality judgment task</td>
<td>1.20 ± 0.67</td>
<td>1.80</td>
<td>.085</td>
</tr>
</tbody>
</table>

*p < .05, **p < .001.

and written lexical richness, oral and written accuracy, and grammaticality judgments, the Time × AO interaction results were significant in favor of the LCLs.

There was no interaction between AO and bilingualism at either measurement, that is, we found the same age effects across bilinguals and monolinguals. Interestingly, however, there was a significant interaction between AO and
Table 2  Impact of age of onset (AO) on English as a foreign language growth

<table>
<thead>
<tr>
<th>Measure</th>
<th>Estimate ± SE</th>
<th>t</th>
<th>Main effect p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receptive vocabulary</td>
<td>8.20 ± 1.21</td>
<td>5.54</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Written lexical richness</td>
<td>0.91 ± 0.27</td>
<td>4.86</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Written fluency</td>
<td>−0.61 ± 0.11</td>
<td>0.87</td>
<td>.386</td>
</tr>
<tr>
<td>Written complexity</td>
<td>−0.08 ± 0.13</td>
<td>−0.07</td>
<td>.786</td>
</tr>
<tr>
<td>Written accuracy</td>
<td>0.33 ± 0.49</td>
<td>2.99</td>
<td>.003**</td>
</tr>
<tr>
<td>Oral lexical richness</td>
<td>0.93 ± 0.37</td>
<td>3.57</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Oral fluency</td>
<td>0.79 ± 0.28</td>
<td>0.35</td>
<td>.722</td>
</tr>
<tr>
<td>Oral complexity</td>
<td>0.01 ± 0.31</td>
<td>0.12</td>
<td>.893</td>
</tr>
<tr>
<td>Oral accuracy</td>
<td>0.77 ± 0.19</td>
<td>2.55</td>
<td>.011*</td>
</tr>
<tr>
<td>Grammaticality judgment task</td>
<td>1.79 ± 2.83</td>
<td>2.74</td>
<td>&lt; .001**</td>
</tr>
</tbody>
</table>

*p < .05, **p < .001.

biliteracy at the end of secondary school as far as half the measures were concerned: AO: receptive vocabulary ($β = 1.74 ± 0.48, t = 3.65$), AO: productive vocabulary ($β = 1.89 ± 1.40, t = 2.35$), AO: written lexical richness ($β = 0.52 ± 0.19, t = 2.71$), AO: written fluency ($β = 1.14 ± 0.76, t = 2.05$), and AO: oral complexity ($β = 0.21 ± 0.11, t = 1.94$). Thus, the early starters of the SIMBI I group not only seemed to profit from their earlier start but also maintained their learning advantage in the long run, outperforming the late starters at T2 with respect to the skills listed, which is reminiscent of age effects in naturalistic settings. Figure 1 illustrates this phenomenon for receptive vocabulary.

Our second research question focused on the mediating influence of contextual factors, that is, the relationship between parental behaviors, attitudes to, and beliefs about EFL and their children’s bilingualism/biliteracy and English learning. It turned out that when other predictors such as bilingualism, biliteracy, and learning environment were added in the models, AO was a relatively weak predictor of FL learning outcome, overshadowed by, for example, large effects of environmental influences, as Figure 2 demonstrates for receptive vocabulary (for all simple effect sizes and likelihood-based $p$ values, see Appendixes S2 and S3 in the Supporting Information online). Number of books at home had a strong and positive impact on 70% of the measures at T1 and 75% of the measures at T2. Very high effect sizes were also found for indirect parental behavior and parental attitudes at both measurements. By contrast, siblings’ attitude predicted success only with receptive vocabulary, written
Figure 1 Receptive vocabulary by age of onset and language group at Time 2.

Turning to the relationship between parental behaviors, attitudes to, and beliefs about EFL and their children’s bilingualism/biliteracy, bilingualism almost always interacted (significantly) with contextual factors at T1: lexical richness (Bilingualism:AttitudesParents $\beta = -0.21 \pm 0.09$, $t = -2.28$), written fluency (Bilingualism:Books $\beta = -0.95 \pm 0.43$, $t = -2.20$, Bilingualism:DirB $\beta = -1.43 \pm 0.32$, $t = -4.49$), written complexity (Bilingualism:AttitudesParents $\beta = -0.07 \pm 0.03$, $t = -2.30$), oral lexical richness (Bilingualism:DirB $\beta = -0.25 \pm 0.13$, $t = -2.03$), oral fluency (Bilingualism:DirB $\beta = -2.58 \pm 1.25$, $t = -2.07$), oral accuracy (Bilingualism:Books $\beta = 0.42 \pm 0.20$, $t = 2.10$), and grammaticality judgments (Bilingualism:DirB $\beta = -0.77 \pm 0.36$, $t = -2.14$); and at T2: listening comprehension (Bilingualism:Books $\beta = -0.76 \pm 0.42$, $t = -1.82$, Bilingualism:DirB $\beta = -0.84 \pm 0.27$, $t = -3.12$), productive vocabulary (Bilingualism:Books $\beta = -2.55 \pm 0.88$, $t = -2.91$, Bilingualism:DirB $\beta = -1.18 \pm 0.54$, $t = -2.17$), receptive vocabulary (Bilingualism:DirB $\beta = -1.22 \pm 0.63$, $t = -1.95$), written lexical richness (Bilingualism:AttitudesParents $\beta = -0.18 \pm 0.08$, $t = -2.19$), written fluency (Bilingualism:Books $\beta = -1.89 \pm 0.47$, $t = -4.03$), written complexity (Bilingualism:AttitudesParents $\beta = -0.09 \pm 0.04$, $t = -2.49$), written accuracy (Bilingualism:Books $\beta = 0.17 \pm 0.06$, $t = 2.68$, Bilingualism:DirB $\beta = 0.13 \pm 0.04$, $t = 3.08$), oral lexical richness (Bilingualism:DirB $\beta = -0.31 \pm 0.10$, $t = -3.28$), oral accuracy (Bilingualism:InB_Parents $\beta = 0.39 \pm 0.14$, $t = 2.81$, Bilingualism:DirB $\beta = 0.26 \pm 0.11$, $t = 2.41$), and grammaticality judgments (Bilingualism:AttitudesParents...
Figure 2 Different types of environmental influences by language group for receptive vocabulary at Time 1.

$\beta = -1.13 \pm 0.31, t = -3.62)$. In other words, bilingualism was effective only in combination with in/direct parental support and positive parental attitudes.

In regard to biliteracy, a similar picture emerged: Biliteracy had a positive influence on about half the measures at both data collection times (see Appendixes S4 and S5 in the Supporting Information online), but there were
always also significant interactions between biliteracy and environmental factors. In other words, biliteracy was always better than monoliteracy, and family involvement/encouragement was always better than no involvement, as the main effects establish. However, biliteracy together with environmental support was particularly effective, compared to either bilingualism or environmental support alone. What is more, the SIMBI I group received substantially more parental support than the other groups, and their parents had markedly more positive attitudes toward FL learning and multilingualism at both measurements.

Finally, it emerges that the random effects made up much of the variance. When clustered data were analyzed without taking the clustering into account (e.g., with analysis of variance [ANOVA]), the Type-I error rate rose to 50%. In other words, if the cluster-randomized design is not considered in cases when participants were not selected at the individual level, variance estimate will always too small, means may be biased, and $p$ values too small, which may suggest age effects that are not really there.

**Discussion**

The main goals of this article were, first, to analyze age effects across different learner populations and, second, to promote further investigations of age of FL acquisition onset within context. A moderate contrast emerged between our four language groups (monolinguals, simultaneous bilinguals, simultaneous bilingual–biliterates, and sequential bilinguals) in terms of AO effects at the beginning of secondary school, when the early starters showed some advantages with respect to a range of productive and receptive L2 measures compared to the late starters. By the end of mandatory school time, however, despite the ECLs’ hugely longer exposure time and earlier AO, these differences had disappeared in all groups save one, namely the simultaneous bilinguals who were also biliterate (SIMBI I group). The question then arises: What is it about the SIMBI I group that allowed their early starters to profit from and maintain benefits from the earlier start?

One early hypothesis was that the SIMBI I children might have profited from the bilingual advantage at the cognitive and linguistic level. However, the significant interactions between bilingualism and contextual aspects suggest that it is the social aspect of bilingualism, especially a supportive learning environment at home, that accounts for cognitive benefits of bilingual experience. Furthermore, like the monolingual group, none of the other bilingual groups (SIMBI II and SEQBI) profited from their earlier AO.
Another hypothesis concerned the biliteracy of the SIMBI I group. As mentioned previously, Bialystok (among others) emphasized greater processing control and superior working memory as fruits of biliteracy resulting in enhanced L3 acquisition. Thus, this group might have also profited from heightened metalinguistic insights resulting from exposure to literacy in two languages (Bialystok, 2007; Sanz, 2000). Because biliteracy is key to success in L3 acquisition, the early biliterates probably profited from this. However, other factors could have also played a role, which brings us to our third hypothesis.

The SIMBI I group received substantial support from their parents, who encouraged biliteracy and were generally interested in FL learning, read to their children, had many books at home, supported multilingualism, and assisted their children with homework. The influence of parental support remained a robust predictor throughout adolescence, and we found similar interactions between context and bilingualism and context and biliteracy at the beginning and end of secondary school. The parents of the SIMBI I students (both early and late) reported more positive attitudes to FL learning and multilingualism than the other parents—hence the significant interaction effects between contextual factors and biliteracy. This would also explain the age effects reminiscent of “earlier = better” findings in naturalistic settings: The intense parental involvement in the case of the SIMBI I group approximates to naturalistic conditions at least regarding affective dimensions and the sense of personal and family relevance with which it must imbue English. Young, naturalistic L2 acquirers are often aware of the role their L2 proficiency plays in securing their well-being (friendships, etc.) and also of family approval for such proficiency because of its usefulness to the family. When parents intensely encourage and support school learning of an additional language, the child must have a similar sense of this learning meeting family goals and of its being integrated into family life.

Finally, it should be mentioned that our analyses profited immensely from the usage of statistical models that (a) take account of clustered data and correlated errors and (b) leave behind grouping assumptions. While ANOVA would have assumed that all our early starters were in one and the same family, class, and school, with the same teachers, parents, and so on, mixed models were able to take into account the oft-neglected cluster effect associated with schools, classes, and families. Almost three decades ago, Willms and Raudenbush (1989) criticized the absence of statistical control over school characteristics as a major limitation for research on school effects. The influence of school context on L2 development is still often overlooked, which is problematic on several levels. While any statistical test will have a 5% chance
of detecting a significant effect even if nothing is going on, this Type-I error rate could easily rise to 50% if clustered data are analyzed without taking the clustering into account.

Conclusion
AO is not only a relatively weak predictor of FL learning outcome compared to contextual factors and bilingualism effects; it is not stable over time and across tasks and skills. It is well known that much of the effect of starting age is the consequence of its covarying relationship with nonbiological factors (Muñoz & Singleton, 2011). It is thus difficult to disentangle maturation from co-occurring contextual, individual, and socio-affective variables; hence, there is an issue with causality. As emerges from our study, age effects are sensitive to contexts and situations. Students under conditions of different school context and climate demonstrate different educational attainment irrespective of AO, which is reflected in the fact that the random effects in our models accounted for much of the variance. Furthermore, not only may different structures show different sensitivity to age of acquisition (e.g., DeKeyser, 2012) but also different tasks/skills (see also Pfenninger, 2017). While spoken and written complexity, fluency, and written accuracy remain relatively unaffected by AO, receptive vocabulary, oral and written lexical richness, oral accuracy, and grammaticality judgments are initially highly sensitive to AO. In the long run, however, none of the tested skills turned out to be problematic as a function of AO, that is, the LCLs displayed faster learning rates than the ECLs.

Turning to the bilingual and biliteracy advantages, not only did the simultaneous bilingual, biliterate children outperform the other groups but they were also differentially affected by AO in terms of benefiting long term from an early start. An earlier AO proved beneficial only for one specific learner group: simultaneous bilinguals who are biliterate and receive substantial parental support (in the short and the long run), as opposed to monolinguals and nonbiliterate bilinguals (simultaneous or sequential).

It is important to mention at this stage how difficult it is to tease apart bilingualism, biliteracy, and environmental influences because they usually interact significantly. As Cenoz (2009) and Aronin and Hufeisen (2009) suggest, we cannot ignore the fact that many factors affect the language acquisition process: Multilingualism is an important but not exclusive influence. Generalizations from past research may have gone further than our current understanding of the phenomenon of the bilingual advantage warrants (Keijzer & Schmid, 2016).

Our results have several research implications. We combined cognitive theory on the age factor with a socio-affective approach to cognitive functioning.
The results, showing that the FL outcomes of multilingual children can be better understood by taking into account a sociolinguistic research perspective, demonstrate the importance of a multidisciplinary research approach to the benefits of bi/multilingualism. There are other relevant issues that could not be accommodated in this article, such as the importance of age-appropriate language pedagogy and classroom practices in primary and secondary school, the development of motivation and attitudes toward learning of EFL as students ascend the educational ladder and its impact on the learning outcome (see Pfenninger & Singleton, 2016), the influence of literacy skills in the L1(s) on literacy development in EFL (see Pfenninger, in press), and learners’ beliefs and opinions related to classroom environmental conditions and the activities they experience (see Pfenninger & Lendl, 2017). Furthermore, what role do teachers’ and peers’ beliefs about, and attitudes toward, early FL instruction play? Are there different age effects for different languages? Do age effects vary across different education programs? Addressing these questions could be fruitful in future studies examining the age factor.

Final revised version accepted 22 May 2018

Notes

1 Time 1: Receptive vocabulary (Biliteracy:InB_Parents $\beta = -1.41 \pm 0.45$), lexical richness (Biliteracy:InB_Parents $\beta = -0.25 \pm 0.07$, $t = -3.63$), written complexity (Biliteracy:InB_Parents $\beta = -0.06 \pm 0.02$, $t = -2.76$, Biliteracy:AttitudesParents $\beta = -0.11 \pm 0.05$, $t = -2.52$), written accuracy (Biliteracy:AttitudesParents $\beta = 0.15 \pm 0.09$, $t = 2.01$), oral complexity (Biliteracy:InB_Parents $\beta = -0.10 \pm 0.03$, $t = -3.03$, Biliteracy:AttitudesParents $\beta = -0.16 \pm 0.06$, $t = -2.46$), oral fluency (Biliteracy:Score1.DirB $\beta = -2.82 \pm 2.01$, $t = 1.41$), oral accuracy (Biliteracy:AttitudesParents $\beta = 0.73 \pm 0.22$, $t = 3.13$). Time 2: Productive vocabulary (Biliteracy:Score.Books $\beta = -3.32 \pm 1.39$, $t = -2.38$, Biliteracy:Score.IndB_Parents $\beta = -1.74 \pm 0.48$, $t = -3.65$), receptive vocabulary (Biliteracy:IndB_Parents $\beta = -1.42 \pm 0.53$, $t = -2.69$, Biliteracy:Score2.DirB $\beta = 1.21 \pm 0.70$, $t = 1.71$), written lexical richness (Biliteracy:Score2.DirB $\beta = 0.16 \pm 0.08$, $t = 1.91$), written complexity (Biliteracy:InB_Parents $\beta = -0.08 \pm 0.03$, $t = -2.42$, Biliteracy:Score2.DirB $\beta = -0.10 \pm 0.04$, $t = -2.57$), written accuracy (Biliteracy:Score2.DirB $\beta = -0.14 \pm 0.06$, $t = -2.94$), oral lexical richness (Biliteracy:InB_Parents $\beta = -0.24 \pm 0.08$, $t = -2.87$), oral fluency (Biliteracy:InB_Parents $\beta = -1.73 \pm 0.86$, $t = -2.01$), oral complexity (Biliteracy:Score.AttitudesParents $\beta = 0.06 \pm 0.04$, $t = 1.32$), oral accuracy (Biliteracy:Score.Books $\beta = 1.23 \pm 0.26$, $t = 4.64$), and grammaticality judgments (Bilingualism: AttitudesParents $\beta = -1.13 \pm 0.31$, $t = -3.62$, Biliteracy:Score2.DirB $\beta = -0.85 \pm 0.32$, $t = -2.68$).
References


Morris, A., Lafontaine, M., Pichette, F., & de Serres, L. (2013). Affective variables, parental involvement and competence among South Korean high school learners of


Supporting Information
Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Appendix S3. Multilevel Regression Analyses for Environmental Influences at Time 2.
Appendix S5. Multilevel Regression Analyses for the Dependent Variables at Time 2.