

# Does younger mean better? Age of onset, learning rate and short-term L2 proficiency in young Danish learners of English

*Teresa Cadierno*

Department of Language and Communication, University of Southern Denmark  
cadierno@sdu.dk

*Mikkel Hansen*

Department of Psychology, University of Paris 8  
mikkel.hansen@iedparis8.net

*Jørgen T. Lauridsen*

Department of Business and Economics, University of Southern Denmark  
jtl@sam.sdu.dk

*Søren W. Eskildsen*

Department of Design and Communication, University of Southern Denmark  
swe@sdu.dk

*Katalin Fenyvesi*

Department of Language and Communication, University of Southern Denmark  
katalin@sdu.dk

*Signe Hannibal Jensen*

Department of Language and Communication, University of Southern Denmark  
sihaje@sdu.dk

*Maria Vanessa aus der Wieschen*

Department of Design and Communication, University of Southern Denmark  
mwies@sdu.dk

## Abstract

This paper reports the results of a semi-longitudinal study investigating the role of age of onset in early foreign language (English) learning. We compared two groups of Danish school children (N = 276) who, following an educational reform in 2014, started their first English classes the same year but at different ages. One group (the early starters) was introduced to English in the 1<sup>st</sup> grade (age 7-8) and the other

group (the late starters) in the 3<sup>rd</sup> grade (age 9-10). Children's receptive vocabulary, receptive grammar, and phonetic discrimination skills were followed for three years, allowing comparisons across groups and time and tracking learning rate and short-term proficiency (after one and two years of instruction). Results showed that the late starters outperformed the early starters in most tests. With respect to learning rate, the tests also revealed that the late starters had an advantage in the receptive grammar test, whereas the phonetic discrimination test showed a more diffuse picture with the late starters seemingly halting in development and the early starters advancing. The results also showed gender differences with boys achieving a higher level of proficiency and exhibiting a faster learning rate than girls. The pedagogical implications of the results are discussed.

**Keywords:** Early foreign language learning; receptive vocabulary; receptive grammar; phonetic discrimination; semi-longitudinal study

## Resumen

Este trabajo presenta los resultados de un estudio semi-longitudinal que investigó el papel de la edad en el aprendizaje del inglés como lengua extranjera. Se compararon dos grupos de niños daneses (N = 276) que, siguiendo una reforma educativa del año 2014, diferían con respecto a la edad en la que comenzaron a recibir enseñanza en inglés. Un grupo comenzó en el primer grado (a la edad de 7-8 años) mientras que el otro grupo empezó en el tercer grado (a la edad de 9-10 años). Se utilizaron tres tests receptivos, uno de vocabulario, otro de gramática y otro que medía habilidades de discriminación fonética. Estos tests se administraron anualmente permitiendo comparar la rapidez de aprendizaje y el grado de aprendizaje de los dos grupos de niños después de uno y dos años de instrucción. Los resultados del estudio mostraron que los niños que empezaron más tarde realizaron la mayor parte de los tests mejor que los niños que empezaron más temprano. En relación a la velocidad de aprendizaje, los tests mostraron que los niños que empezaron más tarde tenían una ventaja con respecto al test gramatical mientras que el test de habilidad de discriminación fonética mostró un resultado más difuso con un parón en el desarrollo para los niños que empezaron más tarde y un mayor avance de los niños que empezaron más temprano. Finalmente, el estudio mostró que los niños alcanzaron un mayor nivel de dominio lingüístico y una mayor velocidad de aprendizaje que las niñas.

**Palabras clave:** Aprendizaje temprano de lenguas extranjeras; vocabulario receptivo; gramática receptiva; discriminación fonética; estudio semi-longitudinal

## 1. Introduction

The issue of the optimal age at which to introduce foreign languages (FL) in schools attracts the attention of educators, policy makers and second language acquisition (SLA) researchers. *The younger, the better* has been a long-standing assumption, i.e., younger children are assumed to be more efficient learners of second languages (L2), and therefore, the younger they start learning an L2, the higher levels of L2 proficiency they will attain (e.g., Hyltenstam, 1992; Krashen et al., 1979). This assumption is in line with educational policies implemented across Europe over the last decades whereby the age at which children start learning FL, predominantly English, in school has dropped (Enever, 2011; Eurydice, 2017). In Denmark, an educational law in 2014 established that children should start learning English from first grade (at age 7 to 8 years) rather than from third grade (9 to 10 years), which had been the status quo in this country since 2002.

Following the earlier introduction of modern languages in most European countries in the last decades, research on the age factor in European classroom settings has increased exponentially. Large-scale investigations have included studies across different European countries (e.g., Enever, 2011) and studies conducted in specific countries such as Spain (e.g., García Mayo & García Lecumberri, 2003 in the Basque country; Muñoz, 2006a, b in Catalonia), Switzerland (e.g., Pfenninger, 2014) and Germany (e.g., Jaekel et al., 2017).

The present study expands the range of geographical contexts that investigate the younger, the better assumption and further contributes with a unique design where we followed two groups of children who started learning English as a foreign language (EFL) within the same year (2014) in Denmark: an early starter group who began English instruction in the first grade (7-8 years old) and a late starter group who began English instruction in the third grade (9-10 years old). English proficiency was measured in three language dimensions, i.e., receptive vocabulary, receptive grammar and phonetic discrimination. Proficiency data for the two age groups were collected at the beginning of their English language instruction, after one, and after two years of instruction. This cross-sectional and short-term-longitudinal design allowed us to examine possible age group differences with respect to rate of learning and short-term proficiency (see also Fenyvesi et al., 2018). It is the first project to investigate the effect of starting age on English FL learning in Danish primary schools since Florander and Jensen (1969) and Mylov (1972).

## **2. Early foreign language teaching in Europe: A policy-based implementation**

The teaching of FLs at younger ages in primary schools around the world has been a common feature of globalized education during the past decades (Enever, 2011). According to Copland et al. (2014), this is due to the above-mentioned assumption of the younger, the better, economic globalization, and parents' demands and expectations which put pressure on governments to introduce English earlier in schools. In Europe, the last decades have witnessed an increased attention to the linguistic and cultural diversity of the European Union (EU), resulting in a European multilingualism policy guided by the Barcelona Council of March 2002, which called for the teaching of two FLs from a very early age.

Recent large-scale studies and reports testify to the implementation of earlier FL programs in Europe. The ELLiE project (Enever, 2011) investigated the teaching of languages in primary schools in seven European countries, i.e., England, Italy, the Netherlands, Poland, Spain, Sweden and Croatia. It found a common trend towards an earlier start in FL education from the first or second year of compulsory schooling. Likewise, the Eurydice (2017) report showed that most European countries have introduced educational reforms to lower the starting age of the first compulsory FL during the last two decades. Generally, students start learning the first FL in primary school before the age of eight although in some countries, e.g., Cyprus and Poland, they even start in pre-primary education.

Given this trend in European primary school systems, a crucial question is whether the younger, the better is a correct assumption as far as instructed FL settings are concerned. This issue will be dealt with in the next section.

## **3. Research into the age factor in naturalistic vs. instructed foreign language settings**

The influence of age on L2 learning has been, and still is, one of the most debated topics in the field of Second Language Acquisition (SLA). According to Muñoz (2008), it is important to differentiate between two distinct strands of research on this issue: one that examines the effect of age of onset (AO) in naturalistic learning settings, i.e., learning through immersion in the second language environment, and another one that investigates this issue in FL settings, i.e., in contexts where learning takes place in a formal context, typically the classroom. This differentiation is needed given the important asymmetries found in both types of contexts regarding the different amount of exposure to the target language (TL) and the different status of ultimate

attainment in the two learning contexts.<sup>1</sup> Whereas amount of exposure to the L2 in naturalistic contexts is in principle unlimited, exposure to the TL in the case of FL learning is generally more restricted, both in terms of the amount of input that the learner is exposed to and the number of interactional opportunities that are afforded. The notion of ultimate attainment also has a different status in the two types of learning situations. Research on age effects on naturalistic second language learning focuses on the final product of the L2 learning process and thus compares the language proficiency attained by early vs. late starters (ES vs. LS) after a minimum of 10 years of residence in the TL country (DeKeyser, 2000). In contrast, studies in FL settings compare ES vs. LS after a given number of instructional hours, which are temporally bounded by the educational system in which the study in question is embedded. Given the relatively limited amount of L2 learning that can take place in instructional settings vis-à-vis naturalistic settings, rate of learning becomes a crucial concern in FL classroom settings (Muñoz, 2008).

Research conducted in naturalistic settings has shown a rate advantage for older learners but an ultimate attainment advantage for younger learners, thus supporting the younger, the better view (e.g., Hyltenstam, 1992; Johnson & Newport, 1989; Snow & Hoefnagel-Höhle, 1978). Studies carried out in FL instructed settings has also shown a rate advantage for the LS over the ES in several language dimensions, e.g., grammaticality judgments (García Mayo, 2003; Pfenninger & Singleton, 2017), degree of foreign accent (García Lecumberri & Gallardo, 2003), production and perception in oral testing (Cenoz, 2003; Muñoz, 2014), composition writing (Cenoz, 2003) and listening and reading comprehension (Jaekel et al., 2017). However, this research has shown that unlike their naturalistic counterparts, younger learners fail to attain higher levels of L2 proficiency in the long term (Muñoz, 2011, 2014).

Two factors that can explain these results are the limited amount of L2 input that learners receive in a typical school context and the fact that the degree of contact with the FL is less intensive (Muñoz & Singleton, 2011; Singleton, 1995). Thus, younger learners would need a much longer period of time to outperform older learners. In addition, the higher cognitive maturity and metalinguistic awareness of older learners may be more adequate in instructional input-limited contexts, where younger learners cannot take advantage of their alleged advantage at implicit learning because of the scarcity of exposure to the TL (DeKeyser, 2000; Muñoz, 2006b; Nikolov & Mihaljević Djigunović, 2006).

#### 4. Effect of age of onset in different L2 proficiency domains

One important finding in previous research on the effects of starting age on L2 proficiency is that it does not affect all aspects of language equally. This has been well documented in studies conducted in naturalistic settings. For example, in a study with English-speaking learners of Dutch in the Netherlands, Snow and Hoefnagel-Höhle (1978) found that whereas older learner groups were especially good at syntactic and morphological acquisition and at metalinguistic ability and vocabulary, differences between the younger and older groups were less pronounced in listening comprehension and story-telling.

In studies conducted in instructed FL settings, age-related differences between the various aspects of language proficiency have also been found. For example, in a study of Swedish-speaking children starting English instruction at Grades 1, 2 and 3, Ekstrand (1982) found a positive correlation between age of onset and performance on tasks such as dictation, reading comprehension, and pronunciation but in more natural communicative tasks (i.e., free oral production), performance was strongly correlated with learners' length of residence in the host country and not with age. Cenoz (2002) found that after six years of English instruction, 11-year-old Basque-Spanish bilinguals did significantly better than 8-year-olds in tasks such as cloze tests and written composition whereas the difference between the two age groups in relation to listening comprehension was only marginal. In the BAF (Barcelona Age Factor) project in Catalonia where 4 different age groups were investigated (AO= 8, 11, 14 and +18), Muñoz (2006b) found that the largest age group differences were observed in the cloze and dictation tests, with an advantage for the late starter groups, whereas a smaller difference between the two age groups was seen in the listening comprehension task (in the initial stages of language acquisition) and in the reception measure on the oral interview (after 726 hours of English instruction).

The development of L2 phonology deserves special mention given its central role in the ongoing debate about the existence of a critical period for L2 learning. According to the Critical Period Hypothesis, this linguistic domain is among the first, if not the only one (Scovel, 1988), to be influenced by a critical period for language acquisition as it not only involves cognitive development but also neuromuscular coordination skills. Research on the acquisition of L2 phonological skills in naturalistic settings has generally supported the younger, the better view showing that even though older learners may outperform younger learners in the initial stages of FL acquisition (e.g., Snow & Hoefnagel-Höhle, 1978), younger learners often end up surpassing older learners (e.g., Krashen et al., 1982; Cook, 1991).

Research in formal instructed settings, on the other hand, has generally provided evidence for a late starting age advantage. For example, in a study examining the development of phonetic skills during six or seven years by children who had started learning English at 4, 8 and 11 years of age, García Lecumberri and Gallardo (2003) found that the older age group was rated as producing less accented and more intelligible FL speech than the two younger groups. As part of the BAF project, Fullana (2006) examined the development of phonetic skills in 4 groups of participants (AO = 8, 11, 14 and 18+ years old) by means of an auditory discrimination task and an imitation task. The results of this study showed that, especially for vowel contrasts, 8-year-olds with 200 and 416 hours of instruction obtained significantly lower correct discrimination scores than the older starting groups. This difference, however, turned out to be non-significant by the time the exposure to English instruction of all the groups had amounted to 726 hours.

The results of the above studies thus show that early starting age is not a factor that facilitates FL sound acquisition in the case of instructed FL learning. In the same vein as other dimensions of L2 proficiency, it may be the case that an early starting advantage requires much longer exposure to and contact with the TL as it is the case in naturalistic contexts (Singleton, 1995).

Age-related differences in relation to various language dimensions have been explained in terms of differences in learners' cognitive maturity. Older learners exhibit higher levels of L2 proficiency in tasks that require higher cognitive maturity such as measures of L2 syntax, morphology and other literacy-related tasks (i.e., vocabulary and reading comprehension) whereas they do not have the same advantage in tests of communicative skills that are considered to be unrelated to the cognitive/academic dimension such as oral fluency and pronunciation (e.g., Cummins, 1980; Muñoz, 2006b). Another explanation that has been offered is related to the presence of implicit vs. explicit mechanisms (Ellis, 1994) that can be at work when engaging in different language activities. Thus, Muñoz (2006b) tentatively concluded that age-related differences are greater in tasks with a higher involvement of explicit processes (e.g., cloze tests) than in those where implicit processes are arguably more present (e.g., oral comprehension task and free production tasks).

## **5. The study**

In the present study, we compared the learning rate and short-term proficiency of two groups of Danish children – an ES group who began English instruction in first grade and a LS group who began English instruction in third grade. Children were given two standardized tests: a receptive vocabulary test (PPTV) and a receptive

grammar test (TROG) in three yearly waves of data collection and in addition a phonetic discrimination task in the second and third years. Part of the data from the first year is outlined briefly in Fenyvesi et al. (2018). Following Unsworth et al. (2014), receptive rather than production tests were chosen as primary language data given previous claims in the literature that children's production during the first two years of instruction is likely to be limited. We also designed and administered a phonetic discrimination task due to the important role that the development of phonological and phonetic skills has had in theoretical discussions and empirical research on the age factor in L2 acquisition (e.g., Flege, 2003; Piske et al., 2001).

The study addressed the following questions:

1. Will ES and LS reach different levels of FL proficiency after one and two years of instruction?
2. Will there be age and gender-related differences in the rate of FL learning during the first two years of L2 English instruction?
3. Will the learning patterns be the same for different dimensions of language proficiency, i.e., receptive vocabulary, receptive grammar, and phonetic discrimination?

## **5.1 Method**

### *5.1.1. Participants*

A total of 276 students (139 boys, 137 girls) participated in the study. All students began formal English instruction in 2014, so in the first year of the study, children were either in first grade (ES aged 7 to 8 years) or in the third grade (LS, aged 9 to 10 years). The children attended 6 elementary schools in the Southern region of Denmark, all located in urban settings (see Cadierno and Eskildsen, 2018 for details). Schools varied with respect to the number of instruction hours per week. In the first year of our study, the schools offered either 1 or 2 weekly English lessons. In the second year, half of the schools offered 1 weekly lesson and the other half 2 weekly lessons to the now 2<sup>nd</sup> graders, and all 4<sup>th</sup> graders received 2 weekly lessons (one lesson is 45 minutes). Given these differences, the number of instruction hours was entered as a factor in the statistical analyses.

The schools followed the guidelines on English language teaching established by the Danish Ministry of Education. The guidelines are phrased in terms of what children



should be able to do after finishing 4<sup>th</sup>, 7<sup>th</sup>, and 9<sup>th</sup> grade. By 4<sup>th</sup> grade, children should be able to participate in short and simple conversations as well as understand and write frequent words, expressions and short texts on everyday topics in English. These learning objectives apply nation-wide, but teachers are free to choose methodology. With young learners, i.e., from grade 1 through 4, the ministry's recommendations for good practice include oral and playful activities, English as the medium of instruction, and teachers' use of gestures and body language to facilitate understanding (see aus der Wieschen, 2017). In terms of educational background, the teachers typically had the required 4-year BA-degree in Education.

We know from classroom observations and recordings that the teachers sing and play with the students and use picture-books and textbooks, the latter mostly with the LS. Language choice practices differ across teachers. Most teachers use both languages to varying degrees. One case study observed a teacher who taught both ES and LS classes. He adhered to an English-only policy with the LS and used a combination of English and Danish with the ES (aus der Wieschen, 2017).

The project was reported to the Danish Data Protection Agency and to the regional ethics committee. School principals had been briefed on the project and agreed to participate. Parents were informed about the project through the schools' intranet and given the option not to let their children participate in the study.

### *5.1.2. Instruments*

As a measure of receptive vocabulary, we used The Peabody Picture Vocabulary Test, Fourth Edition, (Dunn & Dunn, 2007), referred to here as PPVT. As a measure of receptive grammar, we used the Test for Reception of Grammar version 2, (Bishop, 2003), referred to here as TROG. Although not developed for testing EFL, the two tests have been used for that purpose (Dahl & Vulchanova, 2014; Sun et al., 2016; Unsworth et al., 2014). The two tests were administered individually and presented the original English-language items, either a word (PPVT) or a sentence (TROG). Children indicated their answers by selecting one corresponding picture out of four. Items were presented via tape recordings of an expert English speaker<sup>2</sup>. The PPVT has 228 items and exists in two forms, A and B, that were used in alternating years. The administration of the two tests followed the instructions in their respective manuals with the modification that the PPVT was always given starting from the first item, independently of age (as in Unsworth et al., 2014). To allow comparison of scores achieved in PPVT forms A and B, raw scores were converted to Grade Score Values (GSV) following the manual. The maximum GSV score for form A is 270 and for form B 271, both corresponding to a raw score of 228. For the TROG, the same form

with 80 items was used each year. The maximum score was 20 blocks correct (a block is a cluster of 4 items that tests the understanding of a particular linguistic phenomenon) and 80 when scored by individual items correct. We present the results for the TROG in terms of a score of total items passed (e.g., Unsworth et al., 2014). Instructions for both tests were given in Danish.

The phonetic discrimination test consisted of seven minimal pairs, all consisting of monosyllabic words: *free* vs. *three*; *sink* vs. *think*; *these* vs. *ds*; *sue* vs. *zoo*; *log* vs. *lock*; *ice* vs. *eyes*, and *luck* vs. *lock*. The test targets minimal pairs with sounds that are considered to be particularly difficult for Danish learners of English (Mees & Collins, 2000; see Appendix A). The test was administered in blocks of four word pairs in which the minimal pairs were given twice and the single words were repeated in a pair twice along the following lines – but note that the order in which the phenomena occur varied from minimal pair to minimal pair: 1) *free* – *three*; 2) *three* – *three*; 3) *three* – *free*; and 4) *free* – *free*. The children then had to judge whether the two items in the pairs were identical or different. A block was passed when all four pairs in a block were identified correctly. The maximum score was thus 7. For the data analysis, we derived a d-prime score which takes into account that tests where there is no difference in certain pairs may lead to response biases (Macmillan & Creelman, 1991).

### 5.1.3. Procedure

The children were tested individually at their school as part of a multi-day test program. As this was the only year that there was an overlap between the old and the new curricula, practical reasons dictated the start of the school year as the first data collection point. The first testing took place when the first and third-graders started English classes and the subsequent two tests exactly 1 and 2 years later.

The PPVT and TROG tests were administered at all three data collection waves whereas the phonetic discrimination task was administered in waves 2 and 3 (see Table 1). The PPVT was always administered before the TROG within the same data collection session. The phonetic discrimination task (PHON) was administered in a follow up session, typically the day after.

**Table 1.** Data collection points

	Fall 2014 Wave 1	Fall 2015 Wave 2	Fall 2016 Wave 3
Early Starters	Grade 1	Grade 2	Grade 3
Late Starters	Grade 3	Grade 4	Grade 5
Test	PPVT TROG	PPVT TROG PHON	PPVT TROG PHON

#### *5.1.4. Preparation of data for analysis and presentation of model*

We stacked the data for analysis with a mixed effects repeated measures model with the statistics software SAS. As a total of 276 students were included in the study and tested three times, this yielded a potential maximum of  $N = 3 \times 276 = 828$  observations. As some observations were not available for all students, the PPVT had  $N = 802$ , the TROG  $N = 813$ , the PHON  $N = 539$ , and total number of English lessons,  $N = 816$ . For the phonetic discrimination measures, the numbers of observations were lower because they were collected for the two last waves only.

The variables are summarized in Appendix B, and the specification of the model and its advantages is presented in Appendix C.

## **6. Results**

Table 2 reports the scores at waves 1 to 3 for PPVT-GSV, TROG, and the phonetic discrimination task. We analyzed the results with the models specified in Appendix C. Table 3 shows the estimated model for the three outcome measures using a two-way fixed-effects interaction between time and starting grade. The statistical model included the children's classroom as a random effect.

**Table 2.** Raw scores for PPVT-GSV, TROG, and the phonetic discrimination task by starting grade and wave.

		Starting grade			
		1st grade		3rd grade	
	Wave	Mean	Standard Deviation	Mean	Standard Deviation
PPVT	1	80.82	17.07	99.77	20.38
	2	100.28	17.08	115.28	17.07
	3	109.68	23.77	131.57	23.50
TROG	1	8.78	7.69	16.69	9.00
	2	16.12	12.56	27.73	13.42
	3	26.19	15.45	44.02	17.86
PHON	2	2.92	.92	3.80	1.23
	3	3.28	1.09	3.73	1.17

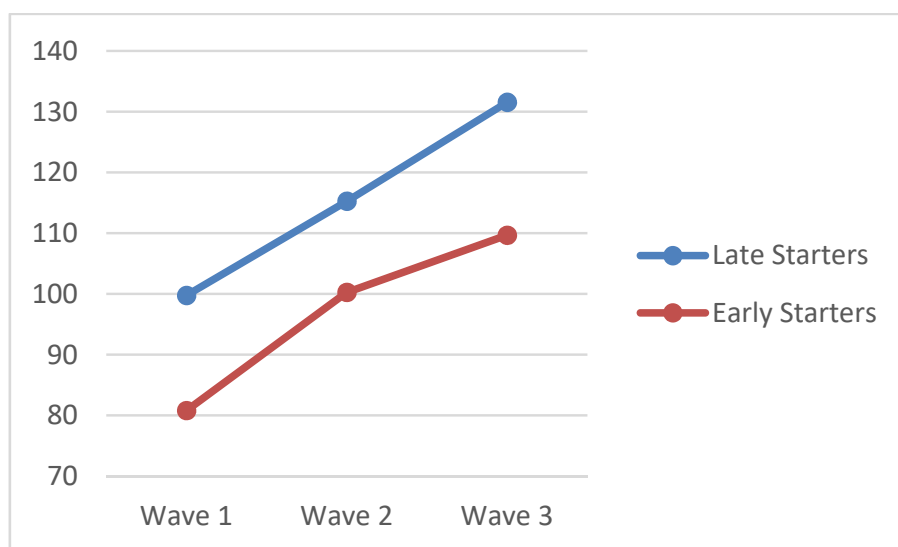
**Table 3.** Estimated two-way fixed-effects regression model of PPVT-GSV, TROG and the phonetic discrimination task.

Variable	PPVT		TROG (items)		PHON (D-prime)*	
	Coefficient	<i>p</i>	Coefficient	<i>P</i>	Coefficient	<i>p</i>
Intercept	94.083	0.000	16.526	0.004	3.703	0.000
Wave2	15.220	0.000	11.202	0.000		
Wave3	32.144	0.000	27.578	0.000	-0.085	0.551
Early Starters (ES)	-17.786	0.000	-7.681	0.000	-0.888	0.000
Wave2*ES	4.123	0.109	-3.791	0.005		
Wave3*ES	-3.236	0.215	-10.200	0.000	0.457	0.013
LessonsTotal	0.035	0.352	0.001	0.979	0.001	0.742

\*) Note that even though the phonetics data were collected only at waves 2 and 3, the results of the first test are here reported on the line “Early Starters.”

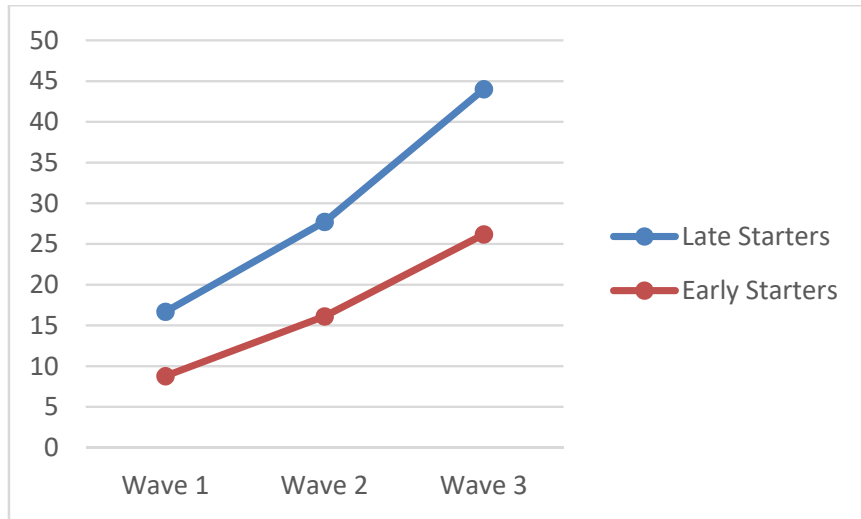
Figures 1 to 3 show the results in actual scores for ES and LS. For the PPVT, Figure 1 shows that ES began with a lower average score than LS, and that the difference between the two groups increased over time. However, while the average difference was significant ( $p < 0.001$ ), neither the increase in difference at wave 2 (the interaction Wave2\*ES,  $p = 0.109$ ) nor wave 3 (Wave3\*ES,  $p = 0.215$ ) were significant. Thus, the difference between ES and LS remained significant at all times of test (Wave1 = 19, Wave2 = 15, Wave3 = 22 points) following the logic of this model. The ES reached the starting level of the LS at wave 2, and at wave 3, they surpassed the ES by about 10 points, the LS having a score at wave 1 of 99.77 and the ES a score at wave 3 of 109.68,  $t(269) = -3.57$ ,  $p < 0.001$ .

Figure 1. PPVT-GSV scores by wave of test and starting grade



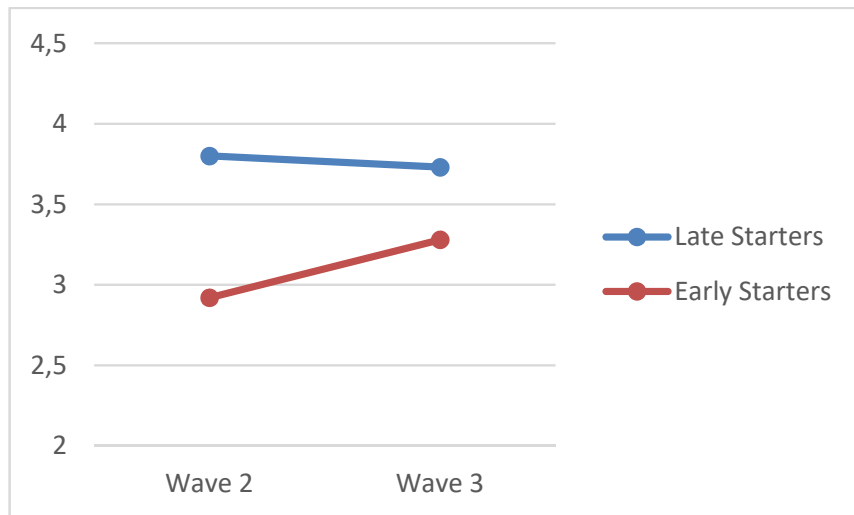
For the TROG, Figure 2 shows a similar development to the PPVT. The difference between ES and LS at wave 1 was significant ( $p < 0.001$ ). However, as opposed to the PPVT, the increases in difference were significant at wave 2 (Wave2\*ES,  $p = 0.005$ ) and wave 3 (Wave3\*ES,  $p < 0.001$ ). Thus, the difference between ES and LS increased significantly with time (Wave1 = 8 points, Wave2 = 12 points, Wave3 = 18 points). At wave 2, the ES had reached the level of the LS at wave 1, and at wave 3 they had surpassed them by about 10 points, the LS having a score at wave 1 of 16.69 and the ES a score at wave 3 of 26.19,  $t(270) = -5.83$ ,  $p < 0.001$ .

**Figure 2.** TROG scores, total items passed, by wave and starting grade



For the phonetics task (d-prime blocks), Figure 3 shows a significant difference between ES and LS at wave 2 (where this test was given for the first time), ( $p < 0.001$ ). However, the difference was significantly reduced at wave 3 to about half a point (Wave3\*ES,  $p = 0.013$ ). There remained, however, a significant difference. The ES did not quite catch up with the LS, scoring 0.44 points below the LS at wave 3, the ES scoring 3.28 at wave 3 compared to the LS scoring 3.73 at wave 3,  $t(256) = 3.09$ ,  $p = 0.002$ .

**Figure 3.** Scores on phonetic discrimination task (d-prime blocks) by wave and starting grade



Finally, as the number of lessons varied between years and classes, we included the total number of lessons of any given student across the span of the study as a predictor in the model. The average total number of lessons for the ES was 132 lessons, corresponding to 99 hours, and for the LS 166 lessons (125 hours). This difference was significant,  $t(200.33) = 10.49$ ,  $p < 0.001$ . Despite the difference, the effect sizes were minimal (see Tables 2 and 3), not even reaching one tenth of a point in any of the models and not reaching significance as a predictor.

In a preliminary analysis of part of the present results, reported in Fenyvesi et al. (2018), we had found an effect of the gender of the students. Therefore, we expanded the model to include gender as a main effect. For all three outcome measures, significant differences were found between girls and boys ( $p$  varying from 0.000 to 0.030), with boys outperforming girls by an estimated 8.5 points on the PPVT-GSV and about 3.5 on the TROG. On the phonetics task, the girls did slightly better than the boys by an average of a quarter of a point. To further understand this effect, we included all interactions with gender and ran the three-way fixed-effects model shown in Table 4.

**Table 4.** Estimated three-way fixed-effects specifications

Variable	PPVT		TROG (items)		PHON (D-prime)	
	Coefficient	p	Coefficient	P	Coefficient	p
Intercept	88.148	0.000	14.806	0.009	3.737	0.000
Wave2	15.304	0.000	9.232	0.000	-	-
Wave3	13.267	0.000	13.412	0.000	0.161	0.412
Early Starters (ES)	-14.131	0.000	-5.466	0.033	-0.804	0.000
Wave2*ES	3.235	0.369	-2.437	0.194	-	-
Wave3*ES	-6.613	0.069	-4.373	0.021	0.317	0.216
Boy	10.216	0.006	3.536	0.143	-0.003	0.989
Wave2*Boy	-0.171	0.965	4.051	0.049	-	-
Wave3*Boy	8.268	0.042	6.851	0.001	-0.530	0.063
ES*boy	-7.377	0.123	-4.473	0.152	-0.177	0.509
Wave2*ES*Boy	1.745	0.733	-2.869	0.279	-	-
Wave3*ES*Boy	-2.697	0.604	-5.086	0.061	0.327	0.370
LessonsTotal	0.041	0.279	0.001	0.980	0.000	0.816

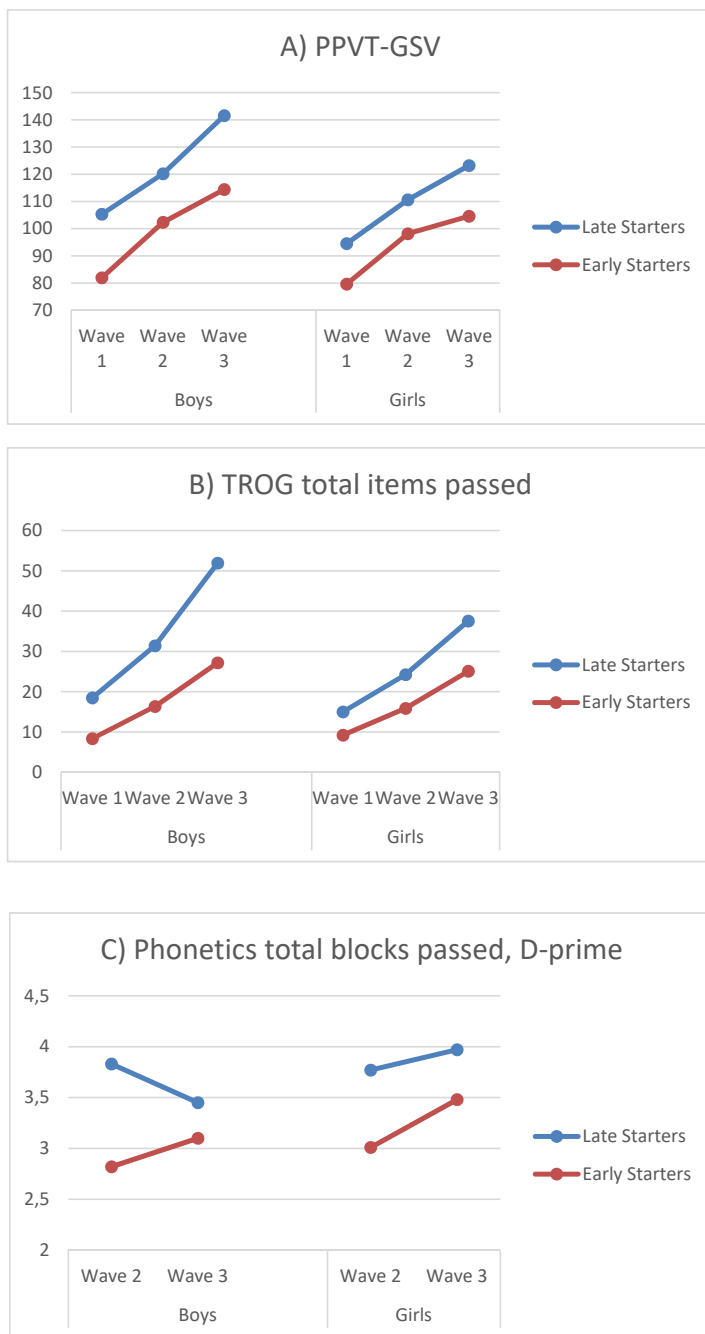
For the PPVT, see Figure 4A, boys obtained higher scores than girls by an estimated 10.22 at the outset of the study,  $p = 0.006$  which did not significantly change at wave 2 ( $p = .965$ ) but at wave 3 the boys did significantly better than the girls, by an estimated difference of 8.27 (Wave3\*boy,  $p = 0.042$ ). However, there were no significant three-way interactions.

For TROG, see Figure 4B, the development curves are quite similar to those from the two-way specification. For ES, the curves for boys and girls are practically coinciding. For LS, the development curve appears more rapid for boys than for girls by about 4 points at wave 2 (Wave2\*boy,  $p = 0.049$ ) and by about 7 points at wave 3 (Wave3\*boy,  $p = 0.001$ ), the so-called rate advantage. There was a tendency toward a statistically significant three-way interaction (Wave3\*ES\*boy,  $p = 0.061$ ), which suggested that LS boys had made bigger gains than girls by wave 3. However, girls also showed a rate advantage which we confirmed by running the analysis from table 2 only for the girls and still found a significant interaction between Wave3\*ES: coefficient 6.80,  $p < 0.001$ .

Figure 4C shows development curves for the phonetics task (d-prime blocks). There was no main effect of gender but a tendency towards a significant interaction (Wave3\*boy,  $p = 0.063$ ), providing a hint that the boys' scores drop a little less than half a point compared to the girls at wave 3.



**Figure 4.** Scores on PPVT (A), TROG (B), and phonetic discrimination task (C) by gender, wave, and starting grade.



## 7. Discussion

A change in Danish educational law provided a unique opportunity to compare the development of English receptive skills by ES and LS who had started L2 English instruction in the same year (2014). The present study examined whether the two groups of learners reached different levels of L2 English proficiency after one and two years of instruction, whether there were age and gender-related differences in the rate of FL learning during the two years of instruction, and whether the learning patterns were the same for the different dimensions of language proficiency. The development of English receptive vocabulary and grammar was tracked across three waves of data collection: wave 1 at the beginning of instruction; wave 2 after one year of instruction; and wave 3 after two years of instruction. The development of English phonetic discrimination skills was tracked at waves 2 and 3, i.e., after one and two years of instruction.

The results of the study showed that in the first wave of data collection, LS obtained significantly higher scores than ES in the two administered language tests. When tested at the beginning of their first year of English instruction, LS were better than ES in relation to receptive vocabulary and grammar. Similarly, when tested at the beginning of their second year of English instruction, LS were also superior in relation to phonetic discrimination skills. This indicates that the starting point for L2 English learning is not the same for both age groups, a finding that in the Danish context is likely due to the accumulated contact with English outside the classroom. This interpretation is supported by the results of previous studies showing that children can and do learn aspects of the L2 *before* English instruction begins (e.g., De Wilde & Eyckmans, 2017; Lefever, 2010; Kuppens, 2010).

Notwithstanding the initial difference between the two age groups, both ES and LS made significant gains in vocabulary and grammar during the first two years of English instruction. Their scores in the receptive vocabulary and grammar tests significantly increased from wave 1 to 2 and from wave 2 to 3. This finding is consistent with the findings reported by Unsworth et al. (2014) who found gains on the same proficiency tests in preschool-aged learners of EFL in the Netherlands. The pattern, however, was different for the phonetic discrimination task as only the ES (but not the LS) made significant gains between waves 2 and 3. The gain of the ES group aside, the LS still outperformed their younger peers which corroborates findings in Fullana (2006) who reported that older starters (11 and 14 years old) tended to discriminate English sounds more accurately than younger starters (8 years old) after 2.5 years of English instruction. The developmental tendency, however, is interesting and at this stage we cannot explain it, partly because we did not administer the phonetic task at wave 1 and so we do not know whether the LS had made gains in phonetic discrimination

during the first year of English instruction and then reached a plateau in relation to this particular test during the second year of instruction. It will be a question for future research to investigate whether phonetic discrimination is an area in which, in a Danish context with a high degree of extramural access to English, it will be an advantage to start early.

Did the initial advantage of LS over ES continue after one and two years of English instruction in terms of receptive vocabulary and grammar? The results of our study showed that for receptive vocabulary, the significant advantage obtained by LS over ES at wave 1 continued across waves 2 and 3. In other words, after 1 and 2 years of English instruction, children starting English classes in the 3rd grade obtained higher scores in receptive vocabulary than children starting English classes in the 1st grade. In relation to receptive grammar, the picture is slightly different. Significant interactions were found between starting grade and scores at waves 2 and 3, meaning that the initial advantage of LS over ES found at wave 1 increased over time, both after the first and second year of English instruction. This finding indicates a rate advantage for LS over ES and supports previous studies which found a rate advantage of LS over ES in grammar-related tasks (e.g., García Mayo, 2003; Jaekel et al., 2017; Florander & Jensen, 1969; Mylov, 1972; Pfenninger & Singleton, 2017). In addition, the fact that LS showed a rate advantage in relation to grammatical skills supports claims made in the literature that LS excel in tasks that require a higher degree of metalinguistic awareness and cognitive maturity, and the use of explicit learning mechanisms at which older students typically excel (DeKeyser & Larson-Hall, 2005; Muñoz, 2006b; Nikolov & Mihaljević Djigunović, 2006).

When comparing the development of the two age groups in the light of the younger, the better debate, for both receptive vocabulary and grammar, the ES group reached the starting level of the LS group after one year of instruction. That is, ES needed one year of English instruction plus potential informal learning to reach the level where LS began without prior formal instruction. In addition, after two years of English instruction, the ES surpassed the point where the LS started by about 10 points – but the LS group still showed a faster learning rate in relation to receptive grammar, which means that the gap between the two groups is widening in favor of the LS. That means that although the ES group did, in fact, do better than their later starting peers at the time of third grade, their advantage is minimal and they are still not, at least not after two years of formal English instruction, catching up with LS with respect to their performance on the PPVT and TROG tests.

Schools differed in whether they offered one or two weekly lessons during the first year of English instruction. In the second year, most schools offered 2 weekly lessons. Results showed that for vocabulary, grammar, and phonetic discrimination,

the total number of lessons over the course of the study was not a significant predictor of proficiency. Unsworth et al. (2014) tested children in a Dutch setting starting at the age of 4 1/2 years. They found that it took on average 1.5 years of instruction before a difference between 60 minutes of English classes or less per week vs 60 minutes or more per week appeared to make a difference in children's results on tests of both receptive vocabulary and grammar. However, in our study, students had participated in two years of English instruction, ranging from 68.25 hours total to 141 hours. Thus, at least in our particular context and age range, an effect of differences in instruction hours has yet to manifest itself. A partial explanation may come from classroom observations and recordings from another part of the larger project which showed that not all time allocated to English classes was necessarily spent on English (see aus der Wieschen and Eskildsen, 2019).

There was a gender gap in receptive vocabulary and grammar scores with boys generally obtaining higher scores than girls. When examining interactions between gender and rate, we noted a tendency towards an advantage in favor of boys as the two years of instruction went by. Boys both obtained higher scores and learned at a more accelerating rate than girls. This was more pronounced for the TROG where this was the case both at waves 2 and 3. A common assumption supported by some studies on young learners in the field of SLA is that females are better at learning FLs than males (e.g., Jaekel et al., 2017; Pae, 2004). Our results go against this assumption and are more in line with the results of Sylvén & Sundqvist (2012) who also found that boys outperformed girls regarding EFL vocabulary. They suggest this may be due to more time spent on gaming, a connection also made by Hannibal Jensen (2017) and Fenyvesi et al. (2018) when relating proficiency scores for a subset of our current sample to the gaming habits of the children. Research into the game-playing habits in young learners has generally shown gender differences regarding gaming frequency – with boys gaming more than girls. According to Sundqvist (2016), this may be due to gender-role stereotyping (i.e., gaming being associated with masculine culture) and lack of female characters in the games that are currently on the market. However, as noted by Carr (2005), female gaming habits may vary worldwide, and gender may thus not be a reliable predictor across the board. In fact, there are several studies conducted with younger learners that have failed to find gender differences in gaming. For example, Butler et al. (2014) found minimal gender differences in the frequency with which Japanese children aged 4 to 12 played various types of digital games. De Wilde & Eyckmans (2017) found that 11-year-old Flemish boys and girls spent a lot of time gaming (more than 1 hour per day), and as result, no gender differences were found in relation to their English proficiency.

L2 motivation is another important factor that may determine the rate and success of L2 attainment (e.g., Dörnyei & Csizér, 1998). Moreover, children's attitudes

towards the FL can be related to their gaming habits. In a recent study that examined the linguistic and attitudinal development of three Swedish children through six years of primary school, Lindgren and Enever (2017) found that the child who was the strongest in terms of English language development was the one who had more access to English at home from the start and who later sought out opportunities to engage in online gaming activities conducted in English with international partners.

Furthermore, children's attitudes and motivation can be also be gender-related. Traditionally it has been found that girls were more motivated to learn FLs than boys (e.g. Heinzmann, 2009), but Oscarson & Apelgren (2005) in a study with Swedish elementary school students found that boys were significantly more motivated than girls to learn English in order to be able to understand and use English media.

## **8. Conclusions, limitations and implications**

This study has compared two groups of Danish school children who differed with respect to age of introduction of English instruction in school and has shown that the LS outperformed their younger peers in most tests employed. While both groups made significant gains on the PPVT and TROG, we saw a rate advantage for the LS on the TROG test. Only the phonetic discrimination test showed a diffuse picture with the LS seemingly halting in development and the ES advancing – a point which calls for further research in this specific area. The test results also generally found gender differences with boys achieving a higher level of proficiency and exhibiting a faster learning rate than girls. The phonetics test was an exception in this respect, with the girls exhibiting a higher growth rate than boys.

These results corroborate those of previous studies in not showing an advantage of introducing English earlier. However, in the specific Danish context, the younger group was the first ever to begin receiving English instruction as of the 1st grade which means that age-appropriate pedagogical tools may not yet be available to the teachers. This is different from other studies (e.g., García-Mayo, 2003; Jaekel et al., 2017; Muñoz, 2006a). At the same time, it is this historical situation that has allowed for the unique design of our study as the only one so far to study the two groups, ES vs. LS, simultaneously.

One limitation of the study is that we have only followed the children over a period of two years. Thus, we can only speak to short-term proficiency and not answer the question of whether, over time, younger learners will catch up with older learners. In future studies, we plan to follow the two groups to the end of primary school (age 16) to investigate the role of starting grade on longer-term English proficiency, testing

oral and written productive skills as well as receptive skills. Previous studies examining longer-term effects in instructed contexts, e.g., in Spain (Muñoz, 2011, 2014) have already shown a LS advantage, e.g., but it remains to be investigated whether this finding will also hold up in contexts such as Denmark where English is more readily accessible outside the classroom.

Finally, our results have pedagogical implications. The different starting points for the two groups with respect to proficiency may have consequences for what teachers are able to achieve in the classroom. As mentioned above, qualitative data point to differences in classroom practices with respect to language choice. Similarly, opportunities for interacting in English may be established at a different level with the LS, for example through a higher degree of student involvement in choosing materials as well as planning and managing activities in the classroom. Because the LS are slightly more proficient in English at the start of instruction (by around 15 to 20 points on the vocabulary and grammar tests), it may at this point in time be somewhat more straightforward for teachers in Denmark to orchestrate meaningful classroom activities with this group. Looked at from the perspective of usage-based approaches to L2 learning (Cadierno & Eskildsen, 2015), this aspect becomes all the more important because of the assumption that meaningful use is the source of learning. Age does play a role in FL teaching, it seems, but younger is not better in this context.

## **Acknowledgements**

We thank the Danish Council for Independent Research, Humanities for funding (DFR4001-00046). Also, we thank school principals, teachers, parents, and research assistants. Finally, and most importantly, we thank the students.

## **Notes**

1. Even though for English, this distinction is somewhat blurred in the Danish context due to the easy access to English in everyday life, the amount of exposure and interaction in this language is presumably not as big as in second language contexts.
2. All three tests presented the test items via tape recordings. Listening to an unfamiliar voice on tape may add a cognitive load that could affect children's linguistic performance differently at different ages (e.g., Field, 2018; Winke et al., 2018). However, on the PPVT and the TROG, results showed that the ES had already at wave 2 caught up with the LS' scores from wave 1 so the

presentation method does not appear to bias the results of the ES. Further, it has been used in previous studies with children of the same age range (e.g., Unsworth et al., 2014), allowing for comparisons across studies.

## 9. References

- aus der Wieschen, M. V. (2017). *Classroom practices in early foreign language teaching in Denmark: on the role of quantity and quality of exposure to English inside the classroom*. Unpublished Ph.D. Thesis: University of Southern Denmark.
- aus der Wieschen, M. V., & Eskildsen, S. W. (2019). Embodied and occasioned learnables and teachables in early EFL classrooms. In H. T. Nguyen & T. M. (Eds.), *Conversation analytic perspectives on English language learning and teaching in global contexts: Constraints and possibilities*, Clevedon: Multilingual Matters, 31-58.
- Cadierno, T., & Eskildsen, S. W. (Eds.). (2015). *Usage-based perspectives on second language learning*. Berlin / New York: Mouton de Gruyter.
- Cadierno, T., & Eskildsen, S. W. (2018). The younger, the better? A usage-based approach to learning and teaching of English in Danish primary schools. *European Journal of Applied Linguistics*, 6(1), 171-182.
- Basbøll, H. (2005). *The phonology of Danish*. Oxford: Oxford University Press.
- Bishop, D. V. M. (2003). *Test for Reception of Grammar (version 2)*. London: Hartcourt Assessment.
- Butler, T. G., Someya, Y., & Fukuhara, E. (2014). Online games for young learners' foreign language leaning. *ELT Journal*, 68(3), 265-275
- Carr, D. (2005). Context, gaming pleasures and gendered preferences. *Stimulation and Gaming*, 36(4), 464-482.
- Cenoz, J. (2002). Age differences in foreign language learning. *International Journal of Applied Linguistic*, 135(1), 125-142.
- Cenoz, J. (2003). The influence of age on the acquisition of English: General proficiency, attitudes and code mixing. In M. del P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 77-93). Clevedon: Multilingual Matters.
- Cook, V. (1991). *Second language learning and language teaching*. London: Edward Arnold.
- Copland, F, Garton, S., & Burns, A. (2014). Challenges in teaching English to young learners: Global perspectives and local realities. *TESOL Quarterly*, 48(4), 738-762.

Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly*, 14(2), 175-187.

Dahl, A., & Vulchanova, M. D. (2014). Naturalistic acquisition in an early language classroom. *Frontiers in Psychology*, 5, 1-9.

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

DeKeyser, R., & Larson-Hall, J. S. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press. 88-108.

De Wilde, V., & Eyckmans, J. (2017). Game on! Young learners' incidental language learning of English prior to instruction. *Studies in Second Language Learning and Teaching*, 7(4), 673-694.

Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research*, 2(3), 203-229.

Dunn, D. M., & Dunn, L. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test: Manual*. Minneapolis: Pearson Assessments.

Ekstrand, L. H. (1982). Age and length of residence as variables related to the adjustment of migrant children, with special reference to second language learning. In S. D. Krashen, R. C. Scarcella & M. H. Long (Eds.), *Child-adult differences in second language acquisition*, University of California: Newbury House Publishers. 123-158.

Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Enever, J. (2011). *Early language learning in Europe (ELLiE)*. London: British Council.

Eurydice. (2017). *Key data on teaching languages at school in Europe*. Brussels.

Fenyvesi, K., Hansen, M., & Cadierno, T. (2018). The role of individual differences in younger vs. older primary school learners of English in Denmark. *International Review of Applied Linguistics in Language Teaching (IRAL)*. <https://doi.org/10.1515/iral-2017-0053>

Field, J. (2018). The cognitive validity of tests of listening and speaking designed for young learners. In S. Papp & S. Rixon (Eds.), *Research and practice in assessing the English of school age young learners* (pp. 128-201). Cambridge, UK: Cambridge University Press.

Flege, J. E. (2003). Assessing constraints on second language segmental production and perception. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in*



language comprehension and production: Differences and similarities (pp. 319-355). Berlin & New York: Mouton de Gruyter.

Florander, J., & Jensen, M. (1969). *Skoleforsøg i engelsk 1959-1965 [School experiment in English classes 1959-1965]*. Stenciled: Denmark's Institute of Education.

Fullana, N. (2006). The development of English (FL) perception and production skills: Starting age and exposure effects. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 41-64). Clevedon: Multilingual Matters.

García Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. del P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters. 115-135.

García Mayo, M. del P. (2003). Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language. In M. del P. García Mayo & M. L. García Lecumberri (Eds.), *Age and the acquisition of English as a foreign language*, Clevedon: Multilingual Matters. 94-114

García Mayo, M. del P., & García Lecumberri, M. L. (Eds.). (2003). *Age and the acquisition of English as a Foreign Language*. Clevedon: Multilingual Matters.

Hannibal Jensen, S. (2017). Gaming as an English language learning resource among young children in Denmark. *CALICO Journal*, 34, 1-19.

Heinzmann, S. (2009). "Girls are better at language learning than boys": Do stereotypic beliefs about language learning contribute to girls' higher motivation to learn English in primary school? *Bulletin Suisse de Linguistique Appliquee*, 89, 19-36.

Hyltenstam, K. (1992). Non-native features of near-native speakers: On the ultimate attainment of childhood L2 learners. *Advances in Psychology*, 83, 351-368.

Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning*, 67(3), 631-664.

Johnson, J. S., & Newport, E. L. (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(1), 60-99.

Krashen, S. D., M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly*, 13(4), 573-582.

Krashen, S. D., Scarcella, R. S., & Long, M. A. (Eds.) (1982). *Child-adult differences in second language acquisition*. Rowley, MA: Newbury House.

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, 35(1), 65-85.

Lefever, S. (2010). English skills of young learners in Iceland: "I started talking English when I was 4 years old. It just bang... just fall into me." Paper presented at Menntakvika Conference, Reykjavik.

Lindgren, E., & Enever, J. (2017). Employing mixed methods for the construction of thick descriptions of early language learning. In J. Enever & E. Lindgren (Eds.), *Early language learning: Complexity and mixed methods* (pp. 201-221). Bristol: Multilingual Matters.

Macmillan, N. A., C. Creelman, D., & Douglas, C. (1991). *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

Mees, I. M., & Collins, B. (2000). *Sound English – A pronunciation guide for speakers of Danish*. Copenhagen: Akademisk Forlag.

Muñoz, C. (Ed.). (2006a). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.

Muñoz, C. (2006b). The effects of age on foreign language learning: The BAF project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters. 1-40.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29(4), 578-596.

Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *International Review of Applied Linguistics in Language Teaching*, 49, 113-133.

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, 44(1), 1-35.

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35(4), 463-482.

Mylov, P. (1972). *Skolereform i engelsk* [The reform of the English curriculum]. Copenhagen: Munksgaard.

Nikolov, M., & Mihaljević Djigunović, J. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics*, 26, 234-260.

Oscarson, M., & Apelgren, B. M. (2005). *Nationella utvärderingen av grundskolan 2003 (NU-03). Engelska. Ämnesrapport till rapport 251* [National evaluation of the primary school 2003 (NU-03). English. Curriculum report for report #251]. Stockholm: Skolverket

- Pae, T.-Il. (2004). Gender effect on reading comprehension with Korean EFL learners. *System*, 32, 265–281.
- Pfenninger, S. E. (2014). The misunderstood variable: Age effects as a function of type of instruction. *Studies in Second Language Learning and Teaching*, 4, 529-56.
- Pfenninger, S. E., & Singleton, D. (2017). *Beyond age effects in instructional L2 learning, revisiting the age factor*. Bristol: Multilingual Matters.
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of perceived foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191-215.
- Scovel, T. (1988). *A time to speak. A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.
- Singleton, D. (1995). A critical look at the Critical Period Hypothesis in second language acquisition research. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 1-29). Clevedon: Multilingual Matters.
- Snow, C. E., & Hoefnagel-Höhle, M. (1978). Age differences in the pronunciation of foreign sounds. *Language and Speech*, 20, 357-365.
- Sun, H., Steinkrauss, R., Tendeiro, J., & De Bot, K. (2016). Individual differences in very young children's English acquisition in China: Internal and external factors. *Bilingualism: Language and Cognition*, 19(3), 550–566.
- Sundqvist, P. (2016). Gaming and young language learners. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning*. New York: Routledge. 446-458.
- Sylvén, L. K., & Sundqvist, P. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL*, 24(3), 302-321.
- Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2014). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, 36(5), 527-548.
- Winke, P., Lee, S., Ahn, J. I., Choi, I., & Cui, Y. (2018). The cognitive validity of child English language tests: What young language learners and their native-speaking peers can reveal. *TESOL Quarterly*, 52(2), 274-303.

## Appendix A.

The consonants targeted in the phonetic tasks are either alien in terms of the Danish phonemic inventory (/θ/ and /ð/) or (if used) not used to distinguish one word from another (/s/ vs. /z/). The phoneme /ð/ exists in Danish but it has a different place and manner of articulation (it is an alveolar approximant rather than the English dental fricative) and never used in word-initial position.

Moreover, Danish has a tendency to neutralize plosives in final position – a process known as lenition – through, e.g., deaspiration (Basbøll, 2005). So the plosive /k/ is aspirated in syllable-initial position and unaspirated in syllable-final position. This means that the final /k/ in a word such as duck is “weaker” in Danish than in English and closer to a /g/, albeit unvoiced. The test targets this potential problem in the minimal pair *log* vs. *lock*. The vowel lengthening in the English *log* does not have an equivalent in Danish.

Another classical problem for Danish EFL learners concerns the [ɰ]/[ɱ] distinction in spite of the abundance of vowels in the Danish phonemic inventory. Therefore, this distinction has also been included in the minimal pair *luck* vs. *lock*.

## Appendix B

Data for the study (N=3×276)

Variable	Definition	Mean	Std	N
PPVT	PPVT score	104.10	25.02	808
TROG (items)	TROG total items correct	21.72	16.71	813
PHON (D-prime)	D-Prime score for PHON*	3.35	1.14	537
WAVE1	Indicator for Wave 1 (pre-test)	0.33	-	828
WAVE2	Joint indicator for Waves 2 and 3	0.33	-	828
WAVE3	Indicator for Wave 3 (second post-test)	0.33	-	828
ES	Indicator (1 for ES, 0 for LS)	0.60	-	828
BOY	Indicator for gender (1 for boys, 0 for girls)	0.50	-	828
LESSONS	Total lessons passed	146.04	35.22	816
CLASS	Identificator for school class	-	-	828
ID	Identificator for child	-	-	828

\*) Only observed for first and second waves

## Appendix C

The design to be applied is a linear regression specification, reading as

where the betas are regression coefficients, and the variables as specified in Table 1. The above specification is for PPVT; similar specifications apply to TROG scores and the phonetic task (D-prime). For the latter, specifications, given that they are only observed for waves 2 and 3, we simply omit indicators and interactions related to WAVE2\_3 and thus use wave 2 as the reference.

The specification has several advantages. First, it allows for separate development curves for ES and LS by adding interaction terms. Second, it allows for inspection of significance of differences between waves and/or starting grade, which we will use and explore on in the subsequent Results section. Third, it allows for control for effects of gender and number of lessons.

We use the joint indicator WAVE2\_3 rather than just an indicator for wave 2 to assess differences between wave 1 and 2 (= the coefficient for WAVE2\_3) as well as the difference between wave 2 and 3 (= the coefficient for WAVE3). For simplicity, in the body of the text we refer to Wave2.

Furthermore, the specification allows for controlling unobserved effects across subgroups by using a hierarchical random effect specification. Here, we use school classes as the first level and students within school classes as the second level. Thus, the full specification becomes

$$\text{PPVT} = \beta_0 + \beta_1 \text{ WAVE23} + \beta_2 \text{ WAVE3} + \beta_3 \text{ ES} + \beta_4 \text{ ES} * \text{ WAVE23} + \beta_5 \text{ ES} * \text{ WAVE3} + \beta_6 * \text{ BOY} + \beta_7 \text{ LESSONS} + e_{\text{CLASS}} + e_{\text{PUPIL-CLASS}} + e$$

By using the estimated coefficients, the average expected score for each two-way subgroup is calculated as shown in the following table:

	Early starters	Late starters
Wave 1	$\beta_0 + \beta_3$	$\beta_0$
Wave 2	$\beta_0 + \beta_1 + \beta_3 + \beta_4$	$\beta_0 + \beta_1$
Wave 3	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5$	$\beta_0 + \beta_1 + \beta_2$

The above specification represents a two-way fixed-effects interaction between waves and ES / LS, which allows for inspecting different development curves for ES /

LS. An extension to a three-way fixed-effect interaction between waves, gender and ES / LS, which further allows for differences across gender in these development curves, is readily specified by adding interactions with gender and reads as

By using the estimated coefficients, the average expected score for each two-way subgroup is easily calculated as shown in the following table:

	Early starters - girls	Late starters - girls	Early starters - boys	Late starters - boys
Wave 1	$\beta_0 + \beta_3$	$\beta_0$	$\beta_0 + \beta_3 + \beta_6 + \beta_9$	$\beta_0 + \beta_6$
Wave 2	$\beta_0 + \beta_1 + \beta_3 + \beta_4$	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_3 + \beta_4 + \beta_6 + \beta_7 + \beta_9 + \beta_{10}$	$\beta_0 + \beta_1 + \beta_6 + \beta_7$
Wave 3	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5$	$\beta_0 + \beta_1 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7 + \beta_8 + \beta_9 + \beta_{10}$	$\beta_0 + \beta_1 + \beta_2 + \beta_6 + \beta_7 + \beta_8$