The role of individual differences in younger vs. older primary school learners of English in Denmark

Fenyvesi, Katalin; Hansen, Mikkel B.; Cadierno, Teresa

Published in:
IRAL - International Review of Applied Linguistics in Language Teaching

DOI:
10.1515/iral-2017-0053

Publication date:
2020

Document version:
Final published version

Citation for published version (APA):

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use
This work is brought to you by the University of Southern Denmark. Unless otherwise specified it has been shared according to the terms for self-archiving. If no other license is stated, these terms apply:

• You may download this work for personal use only.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
• You may freely distribute the URL identifying this open access version.

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim. Please direct all enquiries to puresupport@bib.sdu.dk

Download date: 15. Sep. 2023
Katalin Fenyvesi*, Mikkel B. Hansen and Teresa Cadierno

The role of individual differences in younger vs. older primary school learners of English in Denmark

https://doi.org/10.1515/iral-2017-0053

Abstract: This study investigated 276 Danish first and third graders’ proficiency in English as a foreign language (FL) in relation to several socio-affective factors after their first year of instruction. The results showed that (a) both age groups made similar gains in receptive vocabulary and grammar, suggesting a similar rate of learning in this short time span; (b) the two groups had different socio-affective profiles. Younger learners exhibited lower levels of FL classroom anxiety and more positive English competence beliefs but had a less incremental mindset and relied more on external authorities as a source of motivation. In contrast, older learners relied less on external authorities and showed a more incremental mindset but exhibited higher levels of FL classroom anxiety and less positive English competence beliefs; and (c) only four socio-affective factors had a clear impact on proficiency: English competence beliefs and an incremental mindset had a positive influence while FL classroom anxiety and the influence of external authorities had a negative influence. The pedagogical implications of the different socio-affective strengths and weaknesses that characterize children in the two age groups are discussed.

Keywords: Danish young learners, FL classroom anxiety, EFL competence beliefs, learner’s mindsets, English receptive vocabulary and grammar

1 Introduction

The study of individual differences in second language acquisition (SLA) has become a major area of inquiry in the SLA field (Ellis 2004). One of the most
studied factors is initial age of learning. Leaving aside the theoretical debate as to whether or not a critical period exists for second/foreign (L2) language learners (see Muñoz and Singleton 2011, for a review), empirical research conducted in naturalistic settings has provided evidence for a long-term advantage for younger starters but a rate advantage for older learners (e.g. Krashen et al. 1979).

In contrast, research conducted on foreign language (FL) instructional contexts (i.e. where the L2 is primarily learned in the classroom) has produced different results. Following the pattern observed in naturalistic contexts, younger classroom learners are slower than older learners, but, unlike their naturalistic counterparts, they fail to attain higher levels of L2 proficiency in the long run (e.g. Mayo and María 2003; Muñoz 2006; Nikolov 2009). In other words, research on early starters in FL contexts show that the observed benefits of an early start in naturalistic contexts do not necessarily take place in limited-input FL contexts (Muñoz and Singleton 2011; Courtney et al. 2015).

These findings contradict legislation in several European countries where the starting age of compulsory FL education has been lowered over the past decades (Nikolov and Mihaljević Djigunović 2006). In Denmark, legislation in 2014 lowered the starting grade for English Foreign Language (EFL) instruction from third to first grade. The introduction of the new legislation constitutes the basis for the present study in that the year 2014 allowed a comparison of children at two different first exposure ages in one and the same school year: children who began English instruction in the first grade (i.e. 7 years old) with children who began in the third grade (i.e. 9 years old).

In addition to starting age, there are a number of other individual factors that may influence FL achievement. Crucially, most of this research has focused on adults. Studies on secondary school students are underrepresented and systematic research on primary school students is rare and piecemeal (Boo et al. 2015; Gürsoy and Akin 2013; Mihaljević Djigunović 2015). In the existing literature on young language learners, two types of research can be identified. Whereas some studies have investigated individual learner factors in different age groups without relating them to L2 achievement (e.g. Gürsoy and Akin 2013; Heinzmann 2013), other studies have examined the relation between learner factors and language achievement.

Building on previous research on young learners’ foreign language learning (FLL), the first aim of the present study was to investigate age-related (first vs. third graders) differences in relation to a wide range of individual factors, including learners’ gender, foreign language classroom anxiety (FLCA), EFL competence beliefs (ECB), learners’ motivation and attitudes, and learners’ mindsets, a factor that has not previously been included in research on primary
school children’s FL learning. The second aim of the study was to examine the relationship between age-related learner factors and L2 receptive vocabulary and receptive grammar achievement after one year of instructed EFL learning. To our knowledge, no previous studies have conducted a systematic examination of such a wide range of learner factors in relation to L2 achievement in such young age groups.

2 Review of the literature

Research on individual differences in L2 learning in naturalistic and instructed contexts has revealed great variability in learning outcomes in primary school-age learners (e.g. Courtney et al. 2015; Muñoz 2006; Pfenninger 2014; Unsworth et al. 2014). Next, we review the learner factors that are relevant for the present study.

2.1 Individual factors in young language learners

2.1.1 Foreign language classroom anxiety

Gardner and MacIntyre (1993: 159) defined language anxiety as an “apprehension experienced by the individual in the language class or any situation in which the language is used.” Horwitz et al. (1986) identified FLCA as a distinct phenomenon from other related forms of anxiety such as communication apprehension, test anxiety and fear of negative evaluation.

Few studies have examined FLCA in younger learners. In a large scale mixed-methods study Chan and Wu (2004) investigated FL anxiety among fifth graders (aged 10–11) learning EFL in Taiwan and found that children of this age tended towards high FL anxiety, and that this was caused by factors such as low proficiency, fear of negative evaluation, competition of games, and pressure from parents and peers. Gürsoy and Akin (2013) examined the relationship between age and FL classroom anxiety in 84 Turkish children between 10 and 14 years of age and found that the younger students were less anxious about FL learning than the older students. Finally, Heinzmann (2013) investigated classroom anxiety in young Swiss learners in German-speaking cantons (aged 9–11 years) and found that anxiety was closely related to children’s achievement-related self-concepts, i.e. the students’ perception of their competence in a certain domain. The extent to which anxiety varied across the different age groups was not investigated in this study.
Concerning the relation between gender and anxiety, research has produced mixed results. Heinzmann (2009) found that after 8–9 months of EFL instruction, third grade girls (9 years) in Swiss primary schools were significantly less anxious about making mistakes than boys. Abu-Rabia (2004), however, found that that Israeli female seventh grade EFL students (aged 12–13 years) obtained poorer language proficiency results and had higher anxiety levels than male students. On the other hand, Gürsoy and Akin (2013), in their study of children’s EFL learning in a state primary school in Turkey found no significant difference between the anxiety of girls and boys (aged 10–14 years). The varying results can possibly be explained by the different ages and/or cultural contexts where the studies were conducted.

2.1.2 Learner’s EFL competence beliefs

Learners’ evaluations of their own FL-competence can be considered part of, at least, three constructs used in the SLA-literature on individual factors, namely self-esteem, self-efficacy, and self-concept. These constructs share a common emphasis on an individual’s beliefs about his or her abilities as a person (Valentine and DuBois 2005), but vary in their degree of specificity and the relative importance of cognitive and evaluative aspects of self-beliefs (Mercer 2011). Whereas self-esteem is a global construct focusing on the overall evaluation of one’s value as a person, self-efficacy is more cognitive in nature and tied to individuals’ expectancy beliefs in relation to very specific tasks in specific contexts. Self-concept, on the other hand, is less context-dependent and contains both cognitive and affective elements, being concerned with individuals’ self-perceptions and self-evaluations in a specific domain (Mercer 2011). Given the difficulty involved in establishing a clear-cut separation between the three constructs, we use the term ECB (see also Wigfield et al. 1997) to refer to children’s self-evaluation of their EFL competence in comparison to other school subjects and their classroom peers.

Researchers in the field of child development have found that elementary school children’s view of their own abilities got more realistic around the middle childhood years (e.g. Wigfield et al. 1997), and that their competence-related beliefs (i.e. estimates of how good children themselves are at a given activity) in different school subjects, such as mathematics and sports, became less positive as children got older (Jacobs et al. 2002).

The few studies on young learners’ FL competence beliefs have found parallel results. Concerning children’s estimation of their own English language proficiency, Mihaljević Djigunović (1993) found that early starters (aged 7 or 8 years) had more positive estimations at the beginning of FL instruction but
that these became more objective three years later. In the Early Language Learning in Europe (ELLiE) study, young learners (aged 7–8 years) were interviewed about their conception of their own FL proficiency as compared to their classmates. Many of the younger learners claimed they were faster than their peers at learning the FL but as they grew older, more children claimed to learn as fast as their peers (Mihaljević Djigunović and Lopriore 2011).

Previous research about the relationship between learners’ ECB and gender has found contradictory results. In a Swiss study on children’s motivation and other individual factors, Heinzmann (2009) found that primary school girls in the third grade (aged 9 years) had significantly higher English-competence beliefs (achievement-related self-concept) than boys of the same age. In a later study, however, Heinzmann (2013) found that 9–12-year-old Swiss boys and girls were approximately equally confident in their English skills.

2.1.3 Motivation and attitudes

Motivation has been conceptualized in many different ways within the psychological and the L2 motivation literature. Research on the role of motivation in FL learning can be dated back to the work of the Canadian sociopsychologists Gardner and Lambert (1959) who focused on the distinction between integrative motivation – referring to a positive attitude toward the L2 language community and an emotional identification with its speakers – and instrumental motivation – referring to a more practical usefulness of the L2. In the 1990s insights from self-determination theory and educational psychology came to the front in research on L2 motivation, and additional types of motivation were suggested, such as intrinsic and extrinsic motivation. Intrinsic motivation refers to motivation driven by an enjoyment or interest in an activity whereas extrinsic motivation is based on rewards extrinsic to the activity itself (Noels et al. 2000). Dörnyei (2009) emphasized that the concept of integrative motivation was limited to the immersion context and was insufficient to describe motivation in many language learning environments, such as a classroom context where learners do not have direct contact to speakers of the L2 outside the classroom and in an international learning environment where English plays the role of a lingua franca and it is no longer clear who the speakers of the L2 are (Dörnyei and Kata 2002). Therefore, based on recent development in mainstream psychology, Dörnyei proposed the “L2 Motivational Self System,” which consists of three components: (1) the ideal L2-self, which refers to the learners’ imagined ideal future self as an L2 speaker; (2) the ought to L2-self, which refers to the attributes the learner believes he/she should possess to meet expectations and avoid possible negative outcomes; and
(3) the L2 learning experience, which concerns motivating factors such as the impact of the teacher, the curriculum, the peer group and the experience of success.

Additionally, research on young language learners’ motivation and attitudes has included variables such as attitudes towards FL lessons (e.g. Dörnyei 2010; Enever 2011), attitudes towards the FL (e.g. Enever 2011; Heinzmann 2013; Olshtain et al. 1990), the importance of significant others, such as the parents and the teacher (e.g. Lindgren and Carmen 2013; Mihaljević Djigunović 2012a; M. Nikolov 1999) and the importance of English as a lingua franca (e.g. Enever 2011; Heinzmann 2013; Muñoz 2014).

Longitudinal studies on child and adolescent language learners have typically found that younger children’s attitudes towards the FL and its speakers are more favorable at the outset than 2 to 3 years later (although see Mihaljević Djigunović 2015). This was the case following children from about 7 to 9 years of age (Mihaljević Djigunović and Lopriore 2011); from 9 to 12 years of age (Heinzmann 2013); from 11 to 13, from 13 to 15, and 15 to 17 years of age (Chambers 1999).

A few studies have made cross-sectional comparisons. On the question of enjoyment of FL-class activities, Chambers (1999) found that British 15-year-olds enjoyed writing, reading, and speaking activities in their German lessons significantly less than 13-year-olds, the difference being about 0.4 points on a 5-point-Likert scale. While Mihaljević Djigunović (1993) also found that early starters (aged 6–7 years) liked learning English more than later starters (aged 9–10 years), she also showed that over the span of 3 years of instruction, the early starters maintained a more favorable attitude than the later starters. Burstall (1977) reported similar results from a study with 17,000 English children learning French. When looking at the children who were successful in their efforts to learn French, Burstall found that even though older children learned the FL more efficiently than younger children (the rate advantage), earlier starters (age 8) had the advantage of maintaining a more favorable attitude toward speaking the language than later starters who were first introduced to French at the age of 11 years.

Importantly, several longitudinal studies have shown that the nature of motivation tends to change with age. For example, in an ethnographic study of Hungarian children (age 6–14) learning EFL Nikolov (1999) found that younger learners tended to have more teacher-related motivation but as they got older, they shifted to more utilitarian and instrumental reasons for learning the FL. Similarly, Heinzmann (2013) found that young learners’ intrinsic (i.e. self-determined) motivation decreased with time while lingua franca-related reasons (i.e. reasons for learning English because it is widely spoken around the whole world), gained in importance as age increased.
Concerning the relation between motivation and attitudes on the one hand and gender on the other, there seems to be an agreement that girls have higher motivation and more positive attitudes towards learning foreign languages than boys. For example, Burstall (1977) found that girls had consistently more favorable attitudes towards FL-learning than boys. Heinzmann (2009) found that after 8–9 months of EFL instruction in the third grade of Swiss primary schools, girls were significantly more motivated to learn English than boys. Dörnyei and Kata (2002) also found that 13–14 year old Hungarian girls were significantly more motivated for FLL than boys. Finally, a questionnaire study by Henry (2009) on gender differences in L2-motivation of 12 and 15 year old Swedish students of English found that boys’ ideal L2-selves (items such as “Speaking English is cool”) weakened from the beginning of grade 7 to the end of grade 9 while girls’ ideal selves strengthened over time.

2.1.4 Learners’ mindsets

Language learners have been found to have different beliefs about the role of linguistic ability in FLL (e.g. Horwitz 1999). Similarly, in the field of child development and educational psychology, Dweck’s (2000) socio-cognitive theory has shown that whereas some people believe intelligence to be a fixed entity that people are born with and cannot change much, others believe that intelligence is malleable and can be developed considerably through effort. Mercer and Ryan (2010) have argued that the same distinction applies within the domain of language learning. In a study on the effect of foreign language mindsets on goal-setting and response in challenging situation, Noels and Lou (2015) and Lou and Noels (2016) found that students with a growth FL mindset aimed more at achieving learning goals and showed less helplessness while students with a fixed FL mindset – especially those who had strong perceived language skills – aimed mainly at performing well and were more afraid of failure.

2.2 Individual differences in relation to outcome measures of FL achievement

Previous studies of individual differences in early FL achievement have shown conflicting results (Mihaljević Djigunović 2012b). Using an affective profile score consisting of 13 questions about students’ attitudes toward English and EFL classes, motivation, self-concept and language anxiety, Mihaljević Djigunović (2006) found a significant positive relationship between the affective profile and
productive speaking and writing skills of Croatian students in 8th-graders and 12th-graders. The relationship was stronger for 8th-graders than 12th-graders. For the 12-graders, it was found that the positive correlations between affective profile and EFL results were stronger for speaking than for writing, and stronger for more complex speaking tasks (i.e. argumentative talk) than easier tasks (i.e. answering questions and picture description).

Studies have consistently found a negative relationship between foreign language anxiety and L2 achievement (e.g. Chan and Wu 2004, and see Horwitz 2001, for a summary). Abu-Rabia (2004), examining 12–13-year-old Israeli students using the FLCAS-scale in relation to reading comprehension, creative writing and spelling found significant negative correlations between FLCA and achievement scores on all three aspects of FL proficiency. Concerning gender differences, this study showed that boys had significantly lower FLCA scores and correspondingly higher scores on all three EFL-tests than girls.

Results from the ELLiE study (Mihaljević Djigunović and Lopriore 2011) showed that young primary school children’s (age 7–10) attitudes towards FLL and motivation, as measured by questions about how they felt about learning the FL in general and learning new words in particular, were related to listening comprehension and lexical diversity in oral production. Furthermore, the students with more positive FL competence beliefs, as measured by interview questions on whether they think they learn faster, slower or as fast as other children, were better at listening comprehension and oral production tasks than those with less positive FL competence beliefs.

Regarding learner’s motivation and attitudes, Olshtain et al. (1990) found a significant positive correlation between 11–12-year-old Hebrew-speaking children’s attitudes and motivation and success in EFL, as measured by an achievement test of EFL listening, reading and writing, although motivation and attitudes only explained 5–8% of the variation in EFL-proficiency. Muñoz and Tragant (2001) found that upper primary school and secondary school students with positive attitudes towards EFL obtained significantly higher scores in a global language proficiency test consisting of a cloze test and a dictation. However, they found no significant relations between attitudes and the results of a grammar test and a listening comprehension test.

Finally, Kim and Kim (2014) found that primary school students’ ideal L2-self, measured by imagining and believing oneself to be a competent English user in the future, had a direct influence on perceived EFL-proficiency and self-reported exam-scores.

In relation to learners’ mindsets, a number of studies have explored the role of implicit theories of intelligence on achievement in mathematics and some L1-achievement for the age group 12–13 years (e.g. Blackwell et al. 2007;
Claro et al. 2016) and found that an incremental mindset was related to higher grades. However, to our knowledge there are no previous studies relating an incremental mindset with outcome measures of L2-acquisition.

In sum, previous work on the relationship between individual factors and L2 proficiency shows that they impact children’s FL achievement differentially. FLCA tends to be negatively correlated with FL-achievement whereas ECB tend to be positively correlated. Motivation seems to correlate with FL-achievement as well, depending on the type of motivation and type of language test. The above studies did not include the variable of gender. Most of the previous work on the relation between children’s individual factors and FL achievement has focused on upper primary school and secondary school children. To our knowledge, the present study is the first to simultaneously examine the relation between a wider variety of learner factors and language achievement in such a young FL learner population.

3 Research questions

The present study addressed the following research questions:
1. To what extent does first and third grade children’s proficiency, as measured by their performance in receptive vocabulary and grammar tests, develop after one year of instruction?
2. Are there age- and gender related differences in relation to the following individual factors: learners’ FLCA, ECB, learners’ motivation and attitudes, and learners’ mindsets?
3. Is there a relationship between the above mentioned individual factors and young Danish children’s EFL development as measured by their performance in receptive vocabulary and grammar tests?

4 Method

4.1 Participants

The sample consisted of 276 students (139 boys, 137 girls) who received EFL instruction at six primary schools in Denmark. There were two groups of children, first graders (age = 7–8 years) and third graders (age = 9–10 years), also referred to as early and late starters, that both began English instruction in 2014, the year the educational law was changed.
Data were collected in four Danish public elementary schools which were selected via a stratified sampling technique with geographical location as stratification variable. Two semi-private schools were added to the sample in order to counterbalance the differences in teaching hours offered to the early vs. the late starters at public schools. Following recommendations from the Danish Ministry of Education, the amount of weekly lessons offered in the public schools was one weekly lesson per week (e.g. 45 minutes per week) for the early starters and two weekly lessons (i.e. 90 minutes per week) for the late starters. The two semi-private schools followed the opposite pattern to that of public schools. Differences between public and semi-private schools are relatively small in Denmark compared to many other countries. Both types of schools follow the curriculum guidelines of the Ministry of Education. Eighty to 85% of the funding for semi-private schools is public and fees are relatively modest as a consequence. It is within the economic reach of the majority of parents to send their children to semi-private schools and there is less of a difference between the socio-economic status compared to other countries (Patrinos 2001). There is no tradition in Denmark for extra-curricular English instruction.

4.2 Instruments

4.2.1 Language proficiency tests

To assess children’s English language proficiency development, the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4) (Dunn and Dunn 2007) and the Test for Reception of Grammar, TROG-2 (Bishop 2003) were used as a part of a larger test battery. The PPVT was used to measure receptive vocabulary skills, while the TROG was used to measure receptive grammar skills.¹

 Following previous studies on young language learners (e.g. Dahl and Vulchanova 2014; Sun et al. 2016; Unsworth et al. 2014), receptive tasks were chosen instead of productive tasks given that children’s production was

¹ Recall that the Peabody Picture Vocabulary Test and the Test for Reception of Grammar, TROG-2, were originally developed to test vocabulary and grammar knowledge in L1 populations. When used with L2 learners, however, the TROG test may be measuring more than the knowledge of certain English constructions. It may be considered a test of more general language receptive skills. In order to successfully complete the test, the testee does not only need to comprehend the meaning of the various types of constructions that he/she is presented with (e.g. negation, reversible in and on) but also the specific lexical items, i.e. the nouns and verbs that appear in the sentence stimuli.
expected to be limited at the beginning of English language instruction. Both tests consisted of a series of pictures where children had to select one picture out of four which best corresponded to a word (PPVT) or a sentence (TROG) with potential maximum scores of 228 (raw scores) 271 (graded scores) for the PPVT. For the TROG, the maximum is 20 when scored by total blocks passed and 80 when scored by total items passed. In the PPVT the child saw a picture of a clown, a flower, a bird and a soap bar, and heard the word “bird”, or in the TROG the child saw a picture of a white scarf, a yellow duck, a yellow scarf and a white box and heard the sentence “The scarf is yellow”.

4.2.2 Questionnaire

A questionnaire was constructed to assess the individual factors included in the present study. The factors included in the questionnaire and the sources of the questions are presented in Table 1. The questions were based on questionnaire items previously used with the same age group (Wigfield et al. 1997; Enever 2011; Gunderson et al. 2013) or modified versions of questions previously used with slightly older children (Heinzmann 2013; Dörnyei 2010; Olshtain et al. 1990). Children’s self-concept is undergoing important changes between the ages of 6 and 10 years (see Eccles (1999) for an overview in relation to achievement) with older children tending to be more able to reflect on their performance in relation to themselves and younger children generally being more optimistic about future performances. We therefore included questions that did not require advanced self-reflection, viz. the questions about precursors to Dörnyei’s ideal L2-self. Furthermore, the response format of the questions was adjusted to the cognitive maturity of the children as suggested by Enever (2011).

The questionnaire was administered in Danish and consisted of 39 items. Given that the first graders were not literate at the beginning of the study, the questions were read aloud by the researcher and children marked their answers on a scoring sheet with 5-point Likert scales. Questions on how much children liked something (e.g. To what extent do you like to say something aloud before the whole class in English? and To what extent do you like to listen to music in English when you have English lessons?) were represented by smileys (Enever 2011) while questions about how much they agreed with a statement (“I am afraid of making a mistake when I speak English.” or “It is fun to learn new words in English.”) were represented by dots of increasing size. Children received a brief training session before answering the questionnaire. The questionnaire was pilot tested in two school classes not included in the project. The pilot testing showed that children from both age groups could understand the questions and
Table 1: Individual factors included in children’s questionnaire.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Source</th>
<th>Questions/statements</th>
</tr>
</thead>
</table>
| FL classroom anxiety (FLCA)                 | Heinzmann (2013)              | 1) To what extent do you like to say something aloud before the whole class in English?  
2) I am afraid of making a mistake when I speak English.  
3) I am afraid of giving a wrong answer in the English lessons. |
| EFL-competence belief (ECB)                 | Wigfield et al. (1997)        | 1) How good are you at English?  
2) Are you good at learning something new in English lessons?  
3) If you think about all the children in your class, from the weakest to the best in English, how good do you think you are yourself? |
| Attitudes toward English lessons            | Dörnyei (2010)                | 1) How much are you looking forward to English lessons?  
2) To what extent could you imagine having more English lessons? |
| Attitudes toward different activities in English lessons | Enever (2011)                | 1) To what extent do you like to sing in English when you have English lessons?  
2) To what extent do you like to listen to music in English ...?  
3) To what extent do you like to write in English ...?  
4) To what extent do you like to read in English ...?  
5) To what extent do you like to speak in English ...? |
| Attitudes toward English language           | Olshtain et al. (1990)        | 1) I think it is fun to listen to someone speaking English.  
2) English sounds irritating. (reversed)  
3) It is fun to learn new words in English.  
4) I like to say words in English aloud. |
| Reliance on External authorities            | Dörnyei (2010)                | 1) I would like to learn English because my parents say I have to.  
2) I would like to learn English because the teacher says I have to. |
| English as Lingua Franca                    | Heinzmann (2013)              | 1) I learn English because many people in the world speak English.  
2) I learn English because then I can talk to people from all over the world. |
| Precursors to Ideal L2-self                 | Dörnyei (2010)                | 1) When I grow up, I will surely be good at speaking English.  
2) I think I will be good at reading books in English when I grow up.  
3) I would like to live in an English-speaking country.  
4) I would like to have friends in English-speaking countries. |
were able to appropriately mark the answers on the scoring sheets. A small number of questions that were difficult for children to understand were modified in the final version of the questionnaire.

### 4.3 Procedures

The project was reported to the Danish Data Protection Agency and to the regional ethics committee. The school principals were informed about the purpose of the project and the data collection details. Parents were informed about the project, through the information website of the school. They had the opportunity to actively opt out of the project by responding via the school intranet.

Two language proficiency tests, the Peabody Vocabulary Test (PPVT) and the Test for Reception of Grammar (TROG-2)) were administered twice to the two groups of children, i.e. the first graders and third graders. The first testing took place in the fall of 2014 when children had just begun English classes and the second one year later, in the fall of 2015, when the same children had started in the second and fourth grades respectively. Thus, both grade-levels had two scores, a pretest and a posttest.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Source</th>
<th>Questions/statements</th>
</tr>
</thead>
</table>
| Mindset      | Gunderson et al. (2013) | 1) One will always be as smart as before, no matter what one does. (reversed)  
2) One can become smarter and smarter if one works hard.  
3) Would you choose to solve easy labyrinths to have a lot right? (reversed)  
4) Would you choose to spell easy words to have a lot right? (reversed)  
5) Would you choose to solve difficult labyrinths so you can be better at them?  
6) Would you choose to spell really difficult words so you can be better at spelling?  
7) Do you think it is right that if a person can’t solve a math problem in ten minutes, then the problem is probably too hard for them to do? (reversed)  
8) I know some kids who don’t do well on their schoolwork. Do you think this means that they are not so smart? |
At the time of the first data collection children had received different amounts of English lessons. This was due to the logistics of data collection which prevented all children from being tested at the same time, and the different amounts of English lessons that were offered to the children (1 vs. 2 weekly lessons, see 4.1. above). For this reason, the amount of English lessons was entered as a variable in the statistical analyses of the data. Despite the initial variation in the amount of lessons we made sure that each child received the second administration of the tests exactly one year after the first one.

The items on the PPVT and the TROG-2 were pre-recorded in English to ensure homogeneity in testing. The language tests were given individually to children in the same order of administration, with the PPVT preceding the TROG-2. For the PPVT, Form A was given at the first data collection point and Form B at the second. As the PPVT was used as a proficiency test of L2 vocabulary, the test was given from the beginning to every child, independent of their age.

The questionnaire was administered between the first and the second language proficiency tests during their regular English classes. The questions were read aloud one by one of the researchers, and the children marked the smiley or dot/rectangle of their choice. After filling in the first half of the questionnaire, the classes were given a short break where they sang an English song with the researcher and did some physical exercise. In half of the classes, the first and the second halves of the questionnaire were administered in reversed order to avoid order effects. Answering all the questions took the students approximately 40 minutes on average.

5 Results

5.1 Developmental changes in receptive vocabulary and grammar after one year of instruction

The first research question inquired about children’s progress during the first year of their English instruction on the PPVT and TROG-2 after one year of instruction. As our sample included children from both public and semi-private Danish schools, we conducted independent samples t-tests for all four outcome measures (pretest and posttest for PPVT and TROG-2) comparing the results from semi-private and public schools. As expected, no significant differences were found, and thus we did not examine this factor further.
Figure 1(a) presents an overview of children’s receptive vocabulary development in terms of their starting grade and their progress during their first year of English classes. As described in Section 4.3, children at both grade-levels had two scores, a pretest and a posttest. Their answers were tallied according to the appropriate manuals (Dunn and Dunn 2007; Bishop 2003). Because the PPVT tests at pretest and posttest used different forms, the raw scores were converted into Grade Score Values (GSV) to allow comparison, following the procedure in the PPVT scoring manual.

Overall on the PPVT, the first graders obtained an average score of 80.79 on the pretest and 100.37 on the posttest. The third graders obtained an average score of 99.77 on the pretest and 115.28 on the posttest. We conducted a repeated measures ANOVA with children’s PPVT-GSV scores as a within-subjects variable (pretest vs. posttest) and the following factors as between-subjects variables, Starting Grade (first vs. third), Gender (boys vs. girls), and Weekly lessons of English instruction (1 vs. 2). There was a main effect of time of test, $F(1, 266) = 195.21, p < 0.001, \eta_p^2 = 0.42$, thus the difference between the average of the first and third graders’ pretest scores of 88.48 and the average of their posttest scores of 106.41 was statistically significant. There was also a main effect of starting grade, $F(1, 266) = 76.43, p < 0.001, \eta_p^2 = 0.22$. Thus, regardless of time of test, the first graders’ scores were significantly lower than the third graders’ scores. The interaction between time of test and starting grade was not significant, $F(1, 266) = 1.08, p = 0.30$. The main effect of gender was marginally significant, at $F(1, 266) = 17.18, p < 0.07, \eta_p^2 = 0.06$. However, there was a significant interaction between gender and starting grade, $F(1, 266) = 4.44, p < 0.04, \eta_p^2 = 0.02$. In the first grade, boys’ and girls’ scores averaging across pretest and posttest were about equal, at 92.34 and 88.82, respectively. In the third grade, however, boys scored 112.77 on average, whereas girls scored about ten points lower, at 102.56.

**Figure 1:** A and B - PPVT GSV scores and TROG scores (total items passed) for early starters and late starters at the time of pretest and posttest.
Across the one year from pretest to posttest, children received a total of 40 lessons for the children who received 1 lesson per week (equal to 30 hours) and 80 lessons per week for children who received 2 lessons per week (60 hours). There was no main effect of the number of lessons per week, $F(1, 266) = 2.103$, $p = 0.15$, $\eta^2_p = 0.01$.

However, there was a marginally significant interaction between time of test and the number of lessons per week $F(1, 266) = 3.56$, $p = 0.06$, $\eta^2_p = 0.01$. At pretest, children with one lesson per week regardless of starting grade, obtained a score of 83.78 points and children with two weekly lessons a score of 93.39 points. At posttest, there was little difference, with 1 weekly lesson leading to a score of 104.41 and 2 weekly lessons leading to a score of 108.49. There were no other significant main effects or interactions.

We attribute this pattern to the fact that children could not all be tested at once. Thus, at pretest some classes had had more lessons than others. We confirmed this by tallying the number of lessons before the pretest and correlating them with pretest and posttest PPVT-GSV scores. The number of lessons before the pretest correlated modestly with children’s pretest scores ($r = 0.19$, $p = 0.003$) but not with their posttest scores ($r = 0.08$, $p = 0.18$). Thus, if there was a tendency towards an initial advantage of having more weekly lessons before the pretest, it was not visible one year later.

For the TROG, we used total items passed (e.g. Unsworth et al. 2014) as the principal outcome measure. Scores at pretest and posttest came from the same form. Figure 1(b) presents an overview of children progress on the TROG in total items passed scores in terms of their starting grade. We used the same model as above for a repeated measures ANOVA with TROG scores as the within-subjects variable. However, as the distribution was positively skewed, we used square-root-transformed TROG total items passed scores for the statistical analysis although we report the averages in terms of the total items passed untransformed scores.

Overall, the first graders obtained average scores of 8.78 on the pretest and 16.11 on the posttest. The third graders obtained average scores of 16.69 on the pretest and 27.72 on the posttest. There was a main effect of time of test, $F(1, 268) = 297.97$, $p < 0.001$, $\eta^2_p = 0.53$, thus the difference between the average of the first and third graders’ pretest scores of 11.96 and the average of their posttest scores of 20.79 was statistically significant. There was also a main effect of starting grade, $F(1, 268) = 89.43$, $p < 0.001$, $\eta^2_p = 0.93$. Regardless of time of test, the average of the first graders’ scores was significantly lower than the third graders’ scores. The interaction between time and starting grade was not significant, $F(1, 268) = 1.08$, $p = 0.8$. The main effect of gender was not significant. However, there was a significant interaction between gender and starting grade,
In the first grade, boys’ and girls’ scores were about equal, averaging across pretest and posttest, at 12.35 and 12.55, respectively. In the third grade, however, boys scored 24.94 on average, whereas girls scored about 5 points lower, at 19.62. There were no other significant main effects or interactions, indicating that for receptive grammar, the number of lessons per week did not make a difference, neither at pretest nor at posttest.

5.2 Age- and gender-related differences in individual factors

The second research question aimed at examining age and gender-related differences in relation to individual factors that have been shown to be critical for early FL learning. The individual factors and the scores by starting grade and gender are listed in Table 2 as well as the results of individual ANOVAs with each individual factor as the dependent variable and starting grade and gender as between-subject independent variables. The scores were composite scores of the means of 2 or more questions that students had answered on 5-point Likert scales.

Five of the individual factors showed a significant development with age. FLCA increased significantly between first and third grade, with girls being more anxious than boys regardless of starting grade. There was a significant main effect of starting grade, $F(1, 271) = 8.47, p = 0.004, \eta_p^2 = 0.03$, and a main effect of gender, $F(1, 271) = 11.45, p = 0.001, \eta_p^2 = 0.041$, but no significant interaction, $F(1, 271) = 1.56, p = 0.213$. Boys’ levels of FLCA increased from 2.44 in first grade to 2.64 in third grade, and girls’ levels of FLCA increased from 2.71 in first grade to 3.24 in third grade.

ECB impacts boys and girls differentially at the different grades. For ECB, the results showed a statistically significant main effect of starting grade, $F(1, 270) = 7.69, p = 0.006, \eta_p^2 = 0.028$, and a main effect of gender, $F(1, 270) = 11.13, p = 0.001, \eta_p^2 = 0.04$. These effects were moderated by an interaction between starting grade and gender, $F(1, 270) = 5.91, p = 0.016, \eta_p^2 = 0.021$. This was the only significant interaction in the results presented in this section. In first grade, boys and girls showed equal levels of ECB, at 3.98 and 3.89 respectively. In third grade, however, the boys continued to have high ECB, at 3.94, whereas the girls had substantially lower levels, at 3.39.

Children’s reliance on external authorities, i.e. their parents’ and English teacher’s opinions, decreased significantly with age, from 2.98 in first grade to 2.13 in third grade. This difference reflected a significant main effect of starting grade on the reliance on parents and teachers, $F(1, 271) = 30.95, p = 0.000, \eta_p^2 = 0.102$. There were no other significant effects.
Table 2: Means, standard deviations, and significance levels from GLM of individual factors by starting grade and gender.

<table>
<thead>
<tr>
<th>Factor</th>
<th>First</th>
<th>Third</th>
<th>Results of GLM</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boys</td>
<td>Girls</td>
<td>Boys</td>
<td>Girls</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>FLCA</td>
<td>2.44</td>
<td>1.05</td>
<td>2.71</td>
<td>1.06</td>
</tr>
<tr>
<td>ECB</td>
<td>3.98</td>
<td>0.78</td>
<td>3.89</td>
<td>0.77</td>
</tr>
<tr>
<td>Attitudes to English lessons</td>
<td>3.47</td>
<td>1.29</td>
<td>3.73</td>
<td>1.13</td>
</tr>
<tr>
<td>Attitudes to activities</td>
<td>3.58</td>
<td>0.92</td>
<td>3.72</td>
<td>0.81</td>
</tr>
<tr>
<td>Attitudes to English language</td>
<td>3.83</td>
<td>0.99</td>
<td>3.67</td>
<td>0.77</td>
</tr>
<tr>
<td>Reliance on external authorities</td>
<td>3.05</td>
<td>1.33</td>
<td>2.88</td>
<td>1.31</td>
</tr>
<tr>
<td>(Precursors to) ideal L2 self</td>
<td>3.77</td>
<td>0.84</td>
<td>3.57</td>
<td>0.81</td>
</tr>
<tr>
<td>Importance of English as lingua franca</td>
<td>4.27</td>
<td>1.13</td>
<td>4.03</td>
<td>1.02</td>
</tr>
<tr>
<td>Mindset</td>
<td>3.79</td>
<td>0.62</td>
<td>3.68</td>
<td>0.67</td>
</tr>
</tbody>
</table>
Children’s attitudes toward their English lessons became less positive with age. First graders indicated a mean liking of English lessons of 3.59 whereas third graders indicated a significantly less favorable attitude, at 3.13, $F(1, 270) = 9.39$, $p = 0.002$, $\eta_p^2 = 0.034$. There were no other significant effects.

Two of the individual factors showed no developmental changes but an effect of gender. This was the case for our measures of ideal L2 selves and of attitudes to activities in the English lessons. Children’s ideas about their involvement with English language in the future, the factor ideal L2-self, were basically the same in first and third grade, at 3.67 and 3.75 respectively, $F(1, 270) = 0.88$, $p = 0.350$. However, boys’ scores were higher than girls’ scores, at 3.83 and 3.57 respectively, $F(0, 270) = 8.07$, $p = 0.005$, $\eta_p^2 = 0.029$. A higher score indicated that boys tended to have a stronger projection of themselves as speakers of English language in the future.

For children’s attitudes toward different activities during their English lessons, there was no effect of starting grade, $F(1, 271) = 2.76$, $p = 0.096$. However, there was a significant effect of gender, $F(1, 271) = 4.39$, $p = 0.037$, $\eta_p^2 = 0.03$. Boys indicated less average interest in activities involving English, at 3.49, than girls, 3.74.

The factor “English as lingua franca”, showed no significant change with starting grade, $F(1, 269) = 1.32$, $p = 0.252$, nor a gender difference $F(1, 269) = 2.71$, $p = 0.101$.

Children’s beliefs about whether intelligence is malleable or fixed (the factor Mindset) also showed considerable developmental differences. There was a main effect of starting grade $F(1, 270) = 17.32$, $p < 0.001$, $\eta_p^2 = 0.06$, but no other significant effects. Their average score on the mindset scale was 3.74 in first grade and 4.04 in third grade. A higher score reflected a more incremental mindset.

### 5.3 The relationship between individual factors and change in receptive vocabulary and grammar achievement

The third research question of the study aimed at examining the individual factors presented in Section 5.2 in relation to our language outcome measures, first for the PPVT and then for the TROG, with a mixed effects generalized linear model (Linck and Cunnings 2015). To arrive at the best fitting model, we entered the factors/effects one by one including all interactions. We eliminated non-significant higher-order interactions and then main effects ($p > 0.10$), leaving all lower-order terms of the significant interactions in the model. Table 3 presents the final fitted model for the PPVT-GSV scores.
The model showed a main effect of time of test, but no main effect of starting grade. However, there was a four-way-interaction that involved both time of test and starting grade, so any lower-order effects involving these factors were moderated by the higher-order interaction. The model showed that gender played a role in children’s vocabulary development, with boys scoring higher on the PPVT-GSV than girls by about 4.57 points when all other factors were held constant ($z = -2.94, p = 0.003$). Any interactions with gender were not significant in predicting children’s vocabulary scores and were omitted from the model. Children’s scores also depended on FLCA, but FLCA had differential impacts on gain depending on children’s starting grade and level of ECB, as evidenced by a four-way interaction between FLCA, ECB, time of test and starting grade ($z = 2.06, p = 0.039$). Figure 2 shows the estimated marginal means when we dichotomized ECB at its mean into high and low ECB for the two different starting grades. Looking at first-graders, it is noticeable that the scores at the pretest were remarkably similar whether they had high or low ECB. In other words, whether or not they considered themselves

Table 3: Results of fitted mixed effects generalized linear model: the influence of individual factors on receptive vocabulary scores (PPVT-GSV).

| PPVT-GSV scores                  | Coef. | Robust Std. Err. | Z     | P>|z|  | [95 % Conf. Interval] |
|----------------------------------|-------|------------------|-------|------|-----------------------|
| StartingGrade                    | -27.61| 26.62            | -1.04 | 0.300| -79.79 24.57          |
| Time                             | 36.97 | 10.33            | 3.58  | 0.000| 16.72 57.23           |
| StartingGrade*Time               | 61.69 | 14.74            | 4.19  | 0.000| 32.82 90.58           |
| Lessons per week                 | 6.41  | 2.32             | 2.76  | 0.006| 1.86 10.97            |
| Time*Lessons per week            | -4.16 | 1.87             | -2.22 | 0.026| -7.82 -0.49           |
| Gender                           | -4.57 | 1.55             | -2.94 | 0.003| -7.62 -1.52           |
| FLCA                             | -0.08 | 5.47             | -0.01 | 0.99 | -10.81 10.65          |
| StartingGrade*FLCA               | 8.24  | 8.58             | 0.96  | 0.337| -8.57 25.05           |
| Time*FLCA                        | -8.68 | 3.08             | -2.81 | 0.005| -14.73 -2.64          |
| StartingGrade*Time*FLCA          | -11.89| 4.09             | -2.91 | 0.004| -19.91 -3.87          |
| ECB                              | 1.91  | 4.67             | 0.41  | 0.682| -7.24 11.06           |
| StartingGrade*ECB                | 12.12 | 6.73             | 1.80  | 0.072| -1.07 25.31           |
| Time*ECB                         | -3.29 | 2.56             | -1.29 | 0.198| -8.31 1.73            |
| StartingGrade*Time*ECB           | -13.68| 3.40             | -4.02 | 0.000| -20.34 -7.01          |
| ECB*FLCA                         | 0.23  | 1.39             | 0.17  | 0.866| -2.49 2.96            |
| StartingGrade*ECB*FLCA           | -2.49 | 2.24             | -1.11 | 0.265| -6.87 1.89            |
| Time*ECB*FLCA                    | 1.98  | 0.87             | 2.28  | 0.023| 0.28 3.69             |
| StartingGrade*Time*ECB*FLCA      | 2.10  | 1.02             | 2.06  | 0.039| 0.10 4.10             |
| Reliance on External Authorities | -1.22 | 0.44             | -2.81 | 0.005| -2.08 -0.37           |
| Mindset                          | 3.80  | 1.28             | 2.98  | 0.003| 1.30 6.30             |
| _Constant                        | 60.78 | 17.01            | 3.57  | 0.000| 27.43 94.13           |
competent in English at the pretest, both groups obtained scores from 80 to 90 points. However, compared to first-graders, third-graders were more accurate in assessing their own competences at pretest. Children with high ECB in fact had higher scores, in the 100–110 range, while children with low ECB had scores between 90 and 100 points. When looking at gain, i.e. the distance between the lines for pretest and posttest, all children except third-graders with low ECB, showed gains of between 10 and 20 points during their first year of instruction. However, for children with low ECB in the third grade, their gain depended on their level of FLCA. When these children had low FLCA (equal to 1 or 2 points), they were predicted to make large gains of around thirty points. However, if these children had high FLCA (equal to 4 or 5 points), they were predicted to make almost no gains during one whole year of instruction.

Children’s mindset scores showed a main effect. Thus, the average child would experience an increase of 3.8 points in their vocabulary score for every 1 point their mindset score increased. Children’s reliance on external authorities also made a difference. For every one point that the reliance on external authorities dropped, they gained 1.22 points in vocabulary.

Figure 2: Predictive marginal means of PPVT GSV scores as a function of FLCA and ECB by starting grade and time of test, with 95% confidence intervals.
We then proceeded to fit a full model for the TROG-2 total items passed, which is presented in Table 4. The final model to some degree resembled that for the PPVT.

In fitting the model, it emerged that gender did not have a significant influence on children’s scores on the TROG, so we dropped this factor. There were main effects of both time of test and starting grade, showing that children made gains at both starting grade levels. This analysis also showed an interaction between starting grade and time of test ($z = 3.28, p = 0.001$), suggesting that third-graders made bigger gains than first-graders of about 3.59 points. Recall, however, that in the previously reported ANOVA, this interaction was not significant, so there is reason to question it, as we will discuss in Section 5.4.

FLCA impacted the scores differentially depending on the starting grade of the children, as can be seen in the significant interaction in Table 4, $z = -3.49, p < 0.001$ and depicted in Figure 3. Third graders’ TROG scores were influenced by their FLCA level in the foreign language classroom, with their scores dropping about 2.5 points for every 1 point that their FLCA increased whereas FLCA had relatively little influence on the first graders.

For average children (meaning setting all other factors at their means), children’s ECB showed a main effect, indicating that for each step of increase in ECB, their TROG scores rose by 3.71 points ($z = 4.71, p < 0.001$). Mindset scores predicted children’s receptive grammar scores with each step towards an incremental mindset on the mindset scale corresponding to an increase of about 1.4 points on the TROG. However, this effect was marginally significant in the present analysis ($z = 1.81, p = 0.071$), but see Section 5.4 below. Finally, relying

| TROG Score (total items passed) | Coef. | Robust Std. Err. | Z  | P > |z|   | [95 % Conf. Interval] |
|---------------------------------|-------|-----------------|----|-----|-----|----------------------|
| Starting Grade                  | 17.94 | 3.2             | 5.60 | 0.000 | 11.67 | 24.21               |
| Time                            | 7.46  | 0.84            | 8.93 | 0.000 | 5.82  | 9.1                 |
| Starting Grade*Time             | 3.59  | 1.1             | 3.28 | 0.001 | 1.44  | 5.73                |
| FLCA                            | 1.19  | 0.97            | 1.21 | 0.225 | -0.73 | 3.11                |
| Starting Grade*FLCA             | -3.71 | 1.06            | -3.49 | 0.000 | -5.79 | -1.62               |
| ECB                             | 3.71  | 0.79            | 4.71 | 0.000 | 2.17  | 5.26                |
| Reliance on external authorities| -0.97 | 0.39            | -2.5 | 0.013 | -1.74 | -0.21               |
| Mindset                         | 1.39  | 0.77            | 1.81 | 0.071 | -0.12 | 2.89                |
| _Constant                       | -10.9 | 5.1             | -2.14 | 0.33  | -20.9 | -0.91               |

Table 4: Results of fitted mixed effects generalized linear model: the influence of individual factors on receptive grammar scores when scored by total items passed (TROG-2).
less on external authorities led to an increase in TROG scores. For each 1 point that the reliance on external authorities decreased, children gained about 1 point on the TROG ($z = -2.5$, $p < 0.013$).

5.4 Statistical issues in the observed effects regarding TROG-scores

We now return to the issue of the significant interaction of starting grade and time on the TROG, and the marginally significant main effect of mindset in the above analysis. This is a critical issue because it suggests that children on the TROG make bigger gains in the third than in the first grade. However, we suspected that this interaction might be an artifact of the model. While the distribution of scores for TROG total items passed overall looked acceptable, we noticed that many children had very low scores. This resulted in a somewhat positively skewed distribution. As we did for the initial ANOVAs, we therefore applied a square root transformation to the scores and re-ran our mixed model. This made the interaction between time of test and starting grade insignificant ($z = 1.53$, $p = 0.18$). We wished to confirm the lack of interaction by tallying the scores differently. Rather than looking at the sum of total items passed on the TROG as above, the test may also be scored according to blocks of items passed, as each item belongs to a block of 4 items, according to the TROG-2 manual.

**Figure 3:** Children’s TROG scores, total items passed, as a function of Foreign Language Classroom Anxiety by starting grade, with 95% confidence intervals.
The score then becomes a score of total blocks passed. This way of scoring, however, engendered the problem that many of the younger children received a score of zero and therefore the distribution was still skewed, with no meaningful options for transformations to address the skewness. Instead, we recoded the scores into a binary score of zero or 1, where 1 was equal to 1 or more blocks passed. This allowed us to conduct a mixed-effects logistic regression with the binary score as the repeated-measures target variable with the same fixed effects and their interactions as above. The full model based on binary scores is presented in Table 5. Class was again included as a random effect. We again found a significant effect of time of test and starting grade but no significant interaction ($z=1.24$, $p=0.214$). Thus, for the moment we conclude that there is little indication in the data to suggest a rate advantage for the late starters. All other results retained the same pattern as the model with the total items passed although children’s mindset scores changed from marginally significant to highly significant.

### Table 5: Results of fitted mixed effects logistic regressions model: the influence of individual factors on receptive grammar scores (TROG-2) when scored by total blocks passed and recoded into binary scores.

| TROG Scores (total blocks passed) | Coef. | Robust Std. Err. | Z   | P > |z| | [95% Conf. Interval] |
|----------------------------------|-------|-----------------|-----|-----|---|-------------------|
| Starting Grade                   | 3.29  | 0.71            | 4.66| 0.000|   | 1.91 – 4.68       |
| Time                             | 1.68  | 0.17            | 9.75| 0.000|   | 1.34 – 2.02       |
| Starting Grade*Time              | 0.37  | 0.30            | 1.24| 0.214| -0.22| 0.96             |
| FLCA                             | 0.24  | 0.2             | 1.23| 0.219| -0.14| 0.62             |
| Starting Grade*FLCA              | -0.56 | 0.23            | -2.38| 0.017| -1.02| -0.1             |
| ECB                              | 0.5   | 0.22            | 2.32| 0.020|   | 0.08 – 0.92      |
| Reliance on external authorities | -0.42 | 0.05           | -8.15| 0.000| -0.52| -0.32           |
| Mindset                          | 0.63  | 0.25            | 2.49| 0.013|   | 0.13 – 1.12      |
| _Constant                        | -5.41 | 1.21            | -4.49| 0.000| -7.78| -3.05           |

### 6 Discussion

A change in the legislation in Denmark had presented the unique opportunity to compare early and late starters of EFL in one and the same year. The present study thus investigated the development of English receptive vocabulary and grammar after one year of instruction by first vs. third starting graders in
Denmark. In addition, it examined children’s age- and gender related differences in relation to individual factors such as learners’ FLCA, ECB, learners’ motivation and attitudes, and learners’ mindsets. Finally, we investigated the relationship between these individual factors and young Danish children’s EFL development.

6.1 Developmental changes in receptive vocabulary and grammar after one year of instruction

Both early and late starters came to the FL class with some level of English vocabulary and grammar, which in the Danish context is likely due to informal learning. The late starters began their English classes in third grade roughly at a level corresponding to the results of the early starters after one year of instruction. Thus, the gains of the early starters in one year of instruction, plus potential informal learning during that year, corresponded to the sum of the informal learning that the late starters had achieved prior to formal instruction.

Children on average learned a significant amount of new vocabulary and grammar during the first year of English classes, be they early starters or late starters. This is in line with Unsworth et al. (2014), who found gains on the exact same two proficiency tests in preschool-aged learners of EFL in the Netherlands. The size of the gain both for vocabulary and grammar was roughly equivalent in both grades. Thus, our results at this point provide little evidence of a rate advantage for older learners which is in contrast with earlier studies conducted with young learners in instructional contexts (e.g. Cenoz 2003; Mayo and María 2003; Muñoz 2006). However, a rate advantage may take longer to manifest itself. In fact, as presented in Section 5.4, there is a hint of evidence for a rate advantage for grammar in the late starters when the TROG is tallied by total items passed. However, when tallying the TROG by total blocks passed, the late starters’ advantage disappeared. Future data from the same students will allow us to determine the extent of a rate advantage.

Schools could choose to offer one or two weekly lessons during the first year of English instruction (see Section 4.1). Results showed that at posttest for both vocabulary and grammar, 1 or 2 lessons per week did not make a difference after one year of instruction. This finding is in line with that of Unsworth et al. (2014) who 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. Thus, it may take more than 1 year of instruction
before the number of lessons per week starts to have an effect. Also, informal classroom observations from another part of the larger project revealed that not all time allocated to classes was necessarily spent on English instruction.

6.2 Age- and gender-related differences in individual factors

Age-related differences primarily related to 5 individual factors. In line with previous studies (Gürsoy and Akin 2013; Chan and Wu 2004), late starters exhibited a higher degree of FLCA than early starters. Conversely, ECB decreased with age, as in Mihaljević Djigunović (1993) and (Mihaljević Djigunović and Lopriore 2011). In our results, however, this effect was moderated by an interaction with gender. In fact, boys showed equally high ECB in both starting grades, whereas girls’ ECB showed a significant drop in the third grade. Attitudes towards English lessons also showed a decrease with age, in line with previous research (Chambers 1999; Heinzmann 2013; Mihaljević Djigunović and Lopriore 2011).

On a more positive note, with age, children became less reliant on external authorities (studying English because their parents and teachers say they must), a finding that is in line with Nikolov (1999). Their mindset also tended to become more incremental, i.e. taken together, students became more intrinsically motivated and tended to believe more in the value of effort to improve one’s ability.

Regarding gender differences, our results showed that boys in both grades were less anxious than girls. Boys in the third grade had higher ECB and they tended to a have more positive projection of themselves as future English speakers than girls (Dörnyei 2009). Surprisingly, however, boys also to a lesser extent than girls enjoyed different activities during the English lessons (singing, listening, writing, reading, and speaking in English). The last finding is in line with Dewaele et al. (2016) who found that the women in their study (ranging from 11 to 75 years of age with a mean age of 24) enjoyed activities in FL classes more than men, and Williams et al. (2002) who found the same pattern for British seventh and ninth-graders learning French or German.

It may be the case that boys enjoy activities in English outside the English classroom. Hannibal Jensen (2017) found that children in a subsample from the present study played computer games in English significantly more often than girls. Boys’ frequent engagement with English games, which required them to understand oral and/or written English input in order to successfully complete them, may have made them less anxious (see also Sundqvist and Sylvén (2014) for similar results from adolescents, and more confident about their English language abilities). This, in turn, gives them a familiar scenario when projecting themselves as future users of the language.
6.3 The relationship between individual factors and change in receptive vocabulary and grammar achievement

Previous research with slightly older children, from 5th grade and up (Abu-Rabia 2004; Chan and Wu 2004) has suggested a straightforward negative correlation between FLCA and outcome measures of FLL. However, in young children, our results show that the relationship is much more complex. Looking at vocabulary, FLCA interacted significantly not only with starting grade but also with ECB (Figure 2). Early starters’ level of FLCA had no influence on their vocabulary and grammar scores, be they high or low in ECB. However, for the late starters, the level of ECB made a difference. When students had high ECB, their level of FLCA essentially did not make a difference for their proficiency gain. However, if students had low ECB, FLCA impacted their gain: If in addition to low ECB, students were also highly anxious, there was hardly any vocabulary gain. But if students in addition to low ECB had low FLCA, then they made large gains, of up to 40 points in some cases. However, this result held up for vocabulary but not for grammar, where ECB did not seem to play a moderating role neither in relation to FLCA nor grammar gain. Instead, higher ECB predicted overall higher grammar levels, be it pretest or posttest.

Mihaljević Djigunović and Lopriore (2011) found a positive impact of competence beliefs on different types of tests: listening and oral production tasks. It may be the case that assessing one’s and other students’ level of vocabulary may be different from assessing the level of grammar proficiency. Recall that late starters were relatively accurate in assessing their level of English vocabulary. High ECB by and large corresponded to a high vocabulary score, whereas low ECB corresponded to a low vocabulary score. This component of accuracy in judgement shown by late starters for their own vocabulary level may take longer to manifest itself when it comes to grammar. For receptive grammar, it was also the case that in the third grade, the higher FLCA level, the lower the score, regardless of time of test (see Figure 3).

The influence of FLCA and ECB on our outcome measures was very different from the other individual factors: reliance on external authorities, and learners’ mindset were independent of starting grade and time of test. The less students relied on external authorities to tell them that they must learn English, the better their score regardless of grade, time of test and type of outcome test.

For the mindset scale it was the case that the more incremental the child’s mindset, the better the test scores at both pretest and posttest and both for vocabulary and grammar. Children’s mindsets also became more incremental in the third grade compared to first grade. Other researchers have investigated
mindsets from the point of view of a particular FL mindset (Mercer 2011; Ryan and Mercer 2012; Lou and Noels 2016) in adult FLL. However, in the present research, we used the original mindset scale for young children developed by Dweck and colleagues which in a more general sense tests children’s beliefs about intelligence and ability. Still, these beliefs in children were related to our outcome measures of FLL. To our knowledge, it has not previously been demonstrated that the more incremental one’s mindset, the better one’s FLL acquisition in childhood. It suggests that at least in a Danish context, children who generally believe that effort may improve one’s abilities have an advantage also when learning English as a foreign language. Theoretically, it may contribute to the argument that the ability to learn a foreign language, just like math skills, lends itself particularly well to a conception that it requires a special talent. Having a growth mindset may help overcome this conception.

Our study has found that attitudes towards the English lessons, towards different activities in the English lessons, and towards the English language had little relation to EFL-outcome. This contrasts with the ELLiE-findings (Mihaljević Djigunović and Lopriore 2011) that attitudes are related to listening skills and lexical diversity. Other studies (e.g. Muñoz and Tragant 2001) have also found that positive attitudes towards language had a positive relation to close test and dictation scores. Differences in results may be due to the different language dimensions being tested and the different cultural contexts where studies are conducted. Additionally, our study also found that precursors to ideal L2 self did not predict outcome. This finding contradicts Kim and Kim (2014) where children’s ideal L2 self directly influenced L2 English proficiency. However, L2 proficiency in that study was measured by participants’ self-reported mid-term scores, which is arguably a less reliable measure than the ones used in the present study. Furthermore, our results support Zentner and Renaud’s (2007) claim that stable ideal-self representations do not emerge before adolescence, reaching maturation at age 17, but we also agree with their note that the emergence of ideal-self representations in middle childhood warrants more study.

While gender played a significant role in the results of 4 of the individual factors presented in Section 6.2, it turned out not to be a moderator of the predictor variables when linked to the outcome measures. Rather, in the overall mixed models across starting grade and time, boys were predicted to have a modest but significant 4-point advantage over girls but this was only the case for receptive vocabulary and not grammar. However, three of the individual factors (FLCA, ECB, and precursors to ideal L2 self) favored boys over girls. It may be the case that the added effect of these three factors provide boys with an overall advantage.
7 Conclusion

The results of the present study point to four main findings. First, early and late starters begin English FL instruction with different levels of receptive vocabulary and grammar. For both language dimensions, third graders start off with roughly the same proficiency level as the one achieved by first graders after one year of instruction. Despite this initial difference in proficiency, both age groups make similar gains over one year of instruction, thus suggesting a similar rate of learning for both groups in the short span of one year. Future analyses of proficiency data collected from the same sample of children over time will allow us to determine the extent to which both age groups continue with similar learning rates or whether late starters end up with a rate advantage over early starters (e.g., Mayo and María 2003; Muñoz 2006).

Second, the characterization of early vs. late starters (ages compared were 7–8 with 9–10-year-olds) in terms of a range of individual factors shows that each age group has strengths and weaknesses. The younger age group exhibited lower levels of FLCA and more positive ECB but tended to rely more on external authorities as a source of motivation, and they showed a slightly less incremental mindset. The older age group, on the other hand, relied less on external authorities and showed a more incremental mindset but they also exhibited higher levels of FLCA, they had less positive ECB (especially girls), and they liked English language and English lessons less, though the latter was more predominant for girls than for boys. Our findings are thus in line with Mihaljević Djigunović’s (2014) call for gaining knowledge about possible differences (strengths and weaknesses) between younger vs. older learners in relation to individual factors that are relevant for early language learning.

Third, when examining the relationship between the various individual factors and receptive vocabulary and grammar, only some of the factors that were examined had a clear influence on L2 proficiency. These were FLCA, ECB, learners’ mindset, and the influence of external authorities as a source of motivation. More specifically, the more positive ECB and the more incremental mindset a child has and the lower their reliance on external authorities, the higher the scores obtained in the pretest and the posttest (both for vocabulary and grammar), regardless of age. In addition, higher levels of proficiency in both language dimensions were also obtained by late starters with lower levels of FLCA, although for vocabulary, high ECB had a protective effect against FLCA. Other factors such as children’s liking of the English language and the various activities done in class did not affect their vocabulary and grammar scores. In other words, what children say about their preferences does not prevent them from learning English.
Fourth, one particular factor in our study stands out among the rest. FLCA clearly had a detrimental effect on vocabulary and grammar acquisition in the third grade. This was particularly pronounced for vocabulary proficiency for late starters with low ECB. In contrast, for early starters outcome and FLCA were not related, reflecting perhaps a generally more optimistic outlook at younger ages (Eccles 1999). These results suggest that FLCA may be a determining factor in the discussion of when it is best to start learning a FL. Provided that early starters continue to gain at the same rate as in their first year of instruction, starting at a younger age might give students the advantage of lower levels of FLCA.

These results of the study have pedagogical implications. Primary school teachers could be made aware of the importance of maintaining low-anxiety English classrooms and of developing positive ECB in their pupils. Interventions targeted at developing learners’ incremental mindsets in relation to FL learning could be developed following the spirit of successful interventions in relation to mathematics knowledge (Blackwell et al. 2007; Lanoë et al. 2015).

Finally, the present study has several limitations. In terms of EFL proficiency, we have only included receptive measures, and we have followed gains in receptive vocabulary and grammar for only one year. In future studies we will also include other measures of language proficiency such a production tasks and investigate the development of proficiency in the same children over a longer period of time. Individual learner characteristics are not stable across ages but present different strengths and weaknesses. Thus, future studies should examine how the various individual factors develop over time in the two age groups and how they affect L2 proficiency. Last but not least, we have not included aptitude measures in this study. As language aptitude can be related to growth mindset (Lou and Noels 2016) and to foreign language anxiety (Sparks and Ganschow 1991, 2007), such measures will be included in future studies conducted on the same cohort of children.

Acknowledgements: This study is part of a larger project funded by the Danish Research Council for Independent Research (DFF 4001 00046). We would like to thank Jan Nielsen and Jørgen T. Lauridsen for their help with the statistical analyses of the data, and the reviewers of this manuscript for their insightful comments. We alone are responsible for the contents of the article.

References


Heinzmann, Sybille. 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 appliquée* 89(spring). 19–36.


http://www.informaworld.com/10.1080/0141920120095735