Applied Linguistics Advance Access published June 24, 2014

Applied Linguistics 2014: 1–21
doi:10.1093/applin/amu024

Contrasting Effects of Starting Age and Input on the Oral Performance of Foreign Language Learners

CARMEN MUÑOZ

Department of English and German Philology, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain
E-mail: munoz@ub.edu

The present study focuses on the influence of starting age and input on foreign language learning. In relation to starting age, the study investigates whether early starters in instructional settings achieve the same kind of long-term advantage as learners in naturalistic settings and it complements previous research by using data from oral performance. In relation to input, this study examines and compares the relative impact on learners' oral performance of different input measures: number of years of instruction, number of hours of curricular and extracurricular lessons, number of hours spent abroad in an English-speaking setting, and current contact with the target language. Film-retelling oral narratives from 160 learners of English are analysed in terms of fluency, lexical diversity, and syntactic complexity. Correlational and regression analyses show that input has a stronger association with measures of oral performance than starting age, and that cumulative exposure and, above all, contact with high-quality input are good predictors of learners' oral performance in the foreign language.

INTRODUCTION

The upsurge of interest in research in foreign language (FL) settings in recent decades has resulted from a combination of factors, some belonging to issues and concerns in the field of second language (L2) acquisition, and some to the current widespread implementation of an early start of FL teaching in schools, which expanded the typology of L2 learners and raised new methodological and assessment issues. For over two decades now, the European Union has encouraged early start policies with the aim of protecting linguistic diversity and promoting multilingualism, for reasons of cultural identity, social integration, and economic growth. In practice, however, the early start policies of the majority of European countries have focused on the English language, which is at present the first FL taught in European schools with very few exceptions, as noted in the *First European Survey on Language Competences* (European Commission 2012). English is also the FL taught in primary schools in many parts of the world, in which early start policies forcefully appeared as a consequence of the increasing role of English as lingua franca and its prominent role in global economy. In addition to the above-mentioned reasons, an early

start of FL learning has been uncritically accompanied by expectations of superior L2 outcomes that are modelled by children's language acquisition outcomes in social and school immersion situations, where the amount of exposure is much larger [though still not as large as in first language acquisition where some 10,000 h of exposure are need to attain basic levels of proficiency (Clark 2003)].

The present article focuses on the impact of starting age and of exposure on FL learning settings, that are characterized by limitations in the quantity and quality of the input provided and scant opportunities for engagement with the target language. First, age-related findings are presented showing that the long-term advantage of an early start in a naturalistic setting may not be observed in an FL setting, which highlights the mediating role of learning context. Then, a new study is reported that has examined the respective contribution of starting age and exposure on the L2 oral performance of learners with an average amount of instruction of 15 years, which allows observing the long-term effects of starting age. The aim of the article is to contribute to the understanding of the respective effects of starting age and input in FL learning and, ultimately, to explore the characteristics of input that have a greater impact on learners' learning history.

AGE AND LEARNING CONTEXT

Interest in age as an explanatory factor of differences in L2 acquisition was initiated by Penfield and Roberts (1959) and firmly established with Lenneberg's 1967 proposal of a Critical Period Hypothesis (CPH) for first language acquisition. The maturational account was extended to L2 acquisition by Johnson and Newport (1989) through an oft-cited study of learners with different age of arrival (AOA) in an immigration setting. Johnson and Newport examined 46 Chinese and Korean learners of English in a grammatically judgment test and found that only early learners reached nativelike scores. They also found a linear decline beginning with an AOA of 7 years and ending at around 15–17 years. Learners who had arrived after that age did not present an age-related pattern. This seminal study sets off a line of research that has featured prominently in the field of L2 acquisition up till the present time (e.g. DeKeyser 2000; Birdsong and Molis 2001; Abrahamsson and Hyltenstam 2008; DeKeyser *et al*. 2010; Granena and Long 2013). The main goal of this line of research has been to examine L2 learners' ultimate attainment (UA) as a function of AOA in the L2-speaking country. The shape of this function continues to be a topic of debate (e.g. Munro and Mann 2005; Singleton 2012; Birdsong 2014). Also, a current topic of debate is the ways in which exposure to the target language may be calculated, since the traditional measure of length of residence (LOR), that is, the number of years since arrival in the L2-speaking country is clearly not satisfactory because it does not account for individual differences in exposure/use and orientations in relation

to L2 and L2 speakers (see Moyer 2004, 2009; Flege 2009; Muñoz and Singleton 2011).

Studies examining learners' UA after a long period of residence in an immigration setting have shown the younger starters' UA to be consistently higher than that of older starters and in some cases nativelike (or near-nativelike; see Abrahamsson and Hyltenstam 2008). On the other hand, when early and late arrivals are compared after a shorter period of residence, the latter are observed to have a more rapid pace of learning. In their review of research, Krashen *et al.* (1979) made clear the distinction between the older learners' rate advantage on the one hand and the younger learners' UA advantage on the other. A good illustration is provided by Snow and Hoefnagel-Höhle's 1978 study of a large group of English learners of Dutch in The Netherlands, ranging in age from 3 years to adult. Participants were tested 3–4 times for a period of ~1 year after the first exposure to Dutch. Results showed a clear advantage of teenagers over the other groups, but their advantage over the younger groups diminished with time, and by the end of the study, the youngest group (6–7 years) was approaching the older children (12–15 years) group scores. The older learners were observed to be especially good at syntactic and morphological rule acquisition, and also at metalinguistic ability and vocabulary, which reflected their superior cognitive development.

The very gradual emergence of the early arrivals' advantage has been observed in a few more recent studies. Jia and Fuse (2007) conducted a 5-year longitudinal study of 10 Mandarin-speaking children and adolescents in the USA and observed the timing of the early arrival advantage in six English morphological structures. It was found that only by the end of the 5-year period, younger AOA significantly predicted the average performance on all the structures, though some were not yet fully mastered. The pattern resembles that of phonological acquisition found in several studies (see among others Flege *et al.* 2005). For example, Jia *et al.* (2006) investigated the effects of time on a large sample of Mandarin speakers with a wide range of AOA and found that an early arrival advantage gradually emerged in L2 (English) vowel perception and production after 3–5 years of English immersion.

It has been suggested that a minimum of 5 years of residence (Snow 1983; Munro and Mann 2005) or even 10 years (DeKeyser 2000) may be necessary in CPH studies to methodologically ensure measurement of UA rather than rate, that is, to safely assume asymptote. Along these lines, Long (2007) suggests that after 10 years LOR ceases to be a determinant factor.

However, considering the equivalent period that may be needed in a typical input-limited FL setting to reach similar amounts of input is revealing of the overwhelming difference existing between the two learning settings, as the distribution of the hours of exposure an L2 user receives in a period of 10 years into weekly rations in an FL context vastly surpasses an individual's life expectations.[1] This supposition reveals the inadequacy of adopting the end-state benefits in the arguments in favour of an early start in FL settings. Given that instructed pupils cannot attain nativelikeness because of input limitations

(in both quantity and quality), it may be claimed that in typical FL learning settings amount of exposure never ceases to be a determinant factor, in contrast to Long's suggestion above (Muñoz 2008).

RESEARCH IN FL LEARNING SETTINGS

In contrast to the wealth of studies in an immigration setting, research into the effects of age on L2 learning in an instructed setting has been scarce. However, as DeKeyser (2013: 57) points out, research needs to be conducted in the relevant educational contexts if studies aim to make any educational pronouncements in relation to the benefits of starting an FL at different ages. This section presents empirical evidence from research in FL settings that is relevant for educational applications and sheds light on the complexity of age effects on L2 learning.

Evidence from the neuroscience field has been recently added to the existing behavioural data.[2] Ojima *et al.* (2011) conducted the first study that addresses the effects of starting age and amount of exposure on children's L2/FL learning using neuroimaging data. The focus in this study was children's online processing capacity, specifically, children's semantic processing of spoken English words. Participants were 350 Japanese primary school children who were 6–9 years old at the beginning of the 3-year-long longitudinal study. Their exposure to English was through school lessons, private lessons outside school, home study, and contact with English-speaking acquaintances outside home. Results from behavioural data (a meaning comprehension test of spoken English) showed that children who had started later scored higher on the English test than those who started earlier and had had the same number of hours of exposure. Moreover, irrespective of whether age of onset was controlled for, longer amount of exposure led to higher English scores. Secondly, results from the event-related potential (ERP)[2] datasets showed that later starting age led to larger N400 amplitudes (implying recognition of semantically incongruous words in a picture–word mismatch experimental task) than those shown by earlier starters with same number of hours. Again, irrespective of whether age of onset was controlled for, longer amount of exposure led to larger N400 amplitudes. Another interesting finding concerning exposure from this study is that children who had received more than 800 h of exposure were particularly sensitive to incongruous meanings and that the N400 kept growing at least up to 2,500 h, which suggests to the authors that 'at least a few thousand hours of learning are necessary for the development of even the most fundamental aspect of FL processing such as semantic processing of single words' (Ojima *et al.* 2011: 203).

Recent findings from various instructional settings show a similar pattern. It needs to be noted that in contrast to previous research (e.g Burstall 1975), current studies do not portray situations in which late-starting pupils and early-starting pupils are mixed up at some point in the same classroom, which may result in a levelling-down effect on the early starters. This upsurge

of research has been facilitated by the nation-wide implementation of early start policies in many countries. A case in point is Spain, where a number of studies have been conducted focusing on the comparison of cohorts of pupils with different starting ages, such as the collection of studies in García Mayo and García Lecumberri (2003) and in Muñoz (2006a). Studies in other European countries are also on the increase (e.g. Kalberer 2007; Myles and Mitchell 2012; Unsworth *et al*. 2012), all yielding consistent results showing a rate advantage for the late starters over the early starters after the same number of instruction hours.

The Barcelona Age Factor (BAF) project provides detailed information about the size and characteristics of the older learners' advantage for different language dimensions and after different amounts of exposure, including longitudinal data as well (see Muñoz 2006a). In that research project, data were collected from almost 2,000 Catalan–Spanish bilingual learners of English, distributed in five groups in terms of their starting age: younger than 6, 8, 11, 14, and older than 18 years. Data from the two main groups (those pupils starting at the age of 8 years and at the age of 11 years) were collected after 200, 416, and 726 h of instruction when they were 10;9 and 12;9, 12;9 and 14;9, and 16;9 and 17;9, respectively. The older learners' advantage was greater in the more cognitively demanding tests, and it diminished with time, which was interpreted as reflecting the narrowing of the gap in cognitive maturation between the two groups (Muñoz 2006b).

Although this investigation extended until the younger participants finished secondary education, given the scarcity of input that characterizes typical FL settings, it may be argued that the results missed the alleged long-term benefits that instructed learners may show if age effects were context-independent. In other words, if the assumption is made that early starters in a typical FL learning setting also have a long-term advantage, then it may be expected that young learners will need a much longer period to outperform older learners when input is limited (Singleton 1989).

Long-term effects of an early start

Examining long-term effects of an early start is one of the greatest challenges of age-related research, because participants' learning trajectories may introduce unwanted variation in type and amount of exposure; in addition, there is considerable reliance on participants' adequate recollection of their language learning histories. The very few studies that have looked at long-term benefits of an early start in instructed learners show mixed findings. Larson-Hall (2008) examined 200 Japanese college students who were divided into early starters (who started studying English between ages 3 and 12 years) and had had 1,923 h of exposure to English, and late starters (who started studying English at age 12 or 13 years), with 1,764 h. The study used a phonemic discrimination task, a grammaticality judgment task, and aptitude test, and a background questionnaire. Larson-Hall found only modest effects for an early

start on the grammaticality judgement task but not on the phonemic task that was dependent on total hours of input, that is, when the range of exposure was between 1,600 and 2,200 h (6/8 h per week) but not when exposure was longer (between 2,000 and 4,000 h). When comparing the two groups, the earlier starters scored statistically higher on the phonemic but not the morphosyntactic measure, but again, the advantage was modest. On the other hand, Al-Thubaiti (2010) examined 132 Saudi Arabic college students who were distributed into a group of early or child starters, who had had 1,021 h of exposure to English, and a group of late or teen starters, who had had 819 h. Participants completed a cloze test and a series of UG-motivated performance tasks as well as a questionnaire. Al-Thubaiti did not find any role for age, exposure, or attitudinal variables with the group as a whole, and only a significant relation with general proficiency (cloze test) for those learners who began between the ages of 3 and 6 years, who did so in the home or in a naturalistic immersion setting, that is, not in an input-limited classroom.

A BAF follow-up study investigated the relationship between starting age and long-term FL outcomes in college learners as well (Muñoz 2011). To control as much as possible for variation in learners' exposure, participants completed an extensive written questionnaire about the amount and type of input they had received both in formal and informal contexts, as well as a one-on-one semi-instructed interview that also addressed learners' perceptions of their learning histories. Participants were 159 high-intermediate/advanced learners of English, whose mean length of exposure to the target language since the beginning of instruction was 13.9 years (2,440.5 h). They all had had at least 10 years of exposure. They had begun studying English on average at age 7.8 years (range 2–15.5). The results showed that there was no correlation between starting age and language outcomes (a general proficiency test, a receptive vocabulary test, and a phonetic identification test). Neither were significant differences found when participants were divided into earlier and later starters, but correlations were significant between language outcomes and several measures of input, including length of instruction in years; number of recent curricular and extracurricular lesson hours; number of hours abroad; and current contact with the target language. In contrast, measures of input that were related to exposure during primary and secondary education did not show significant associations with language outcomes. In brief, amount of exposure over the whole language learning trajectory and at the time of testing, as well as native-speaker input were the input measures that appeared associated to L2 long-term achievement in these learners. In addition, these quantitative results about the significance of input in learners' achievement were later corroborated by the analysis of the participants' interviews. These showed that when the participants identified a turning point in their learning histories, the majority (85.7 per cent) pointed to intensive exposure experiences which had opened the gate to a new productive and meaningful learning phase (Muñoz 2012b).

In sum, research has shown that the long-term advantage conferred to learners by an early start in a naturalistic language learning context may not be found in an FL learning context because of input limitations. Furthermore, recent research findings indicate that amount and type of input may play a more crucial role than starting age in learners' long-term language attainment.

THE STUDY

The present study contributes to the line of research that investigates long-term effects of age by providing data from learners' oral performance, which none of the studies reviewed above have examined. By examining learners' oral performance and contrasting the results with those obtained in previous research, the present study also explores the issue of whether in previous research older starters may have been given an advantage by off-line tasks where they could use declarative knowledge and metalinguistic awareness (Long 2013: 267). This study also aims at enriching our knowledge of the effects of input in long-term FL learning by exploring which input measures (e.g. length of instruction, exposure abroad) have a stronger predictive power in FL learning.

The decision on which measures to use in the present study was informed by previous research findings from the areas of L2 learning and child bilingualism. The rationale for using cumulative measures of exposure such as number of years of instruction and number of hours of curricular and extracurricular lessons is that input quantity is a crucial factor in typical FL learning situations, where input is limited, as seen above. A parallel can be seen in bilingual development where children may be exposed to less input in each of the languages resulting in rate differences with monolingual development (Gathercole and Hoff 2007), which have been observed to disappear when bilinguals and monolinguals are matched on cumulative length of exposure[3] (Unsworth 2013a). Amount of exposure is also an important factor in the explanation of the asymmetries found in the rate of development of the dominant and the non-dominant language of bilinguals (Gathercole 2007; Gathercole and Thomas 2009; Paradis 2010; Hoff *et al.* 2012), although not all aspects of bilingual children's development seem to be similarly affected by the amount of language to which they are exposed (Unsworth *et al.* in press). However, as in naturalistic language acquisition, length of exposure may hide an enormous amount of variation among learners, which advises the use of more detailed measures (Flege 2009). One such measure in both immigrant studies and child bilingualism studies is the amount of native-speaker contact. In the area of speech acquisition research, it has been argued that LOR may be less crucial than access to input from native speakers (Flege and Liu 2001; Flege 2009). In bilingual studies, the amount of native-speaker contact has also been proved to be a predictive factor (Paradis 2011). Although a quantitative measure of native-speaker contact cannot be provided, the present study will use the measure of amount of time spent in an English-speaking context,

which appeared significantly correlated with English proficiency in the previous study, and which may be seen to reflect native-speaker contact as well.

Native-speaker contact has been regarded as a crucial indicator of more qualitative measures of input, such as the linguistic richness of the environment (the amount of native-speaker contact that the children experience via media, playmates, and organized extracurricular activities) (Paradis 2011; see also Jia and Fuse 2007; Unsworth 2013b). In the present study, the linguistic richness of the environment is partly reflected in the measure of current informal contact with the target language, which combines this qualitative dimension with a quantitative dimension of current amount of exposure, which has also been found to be a significant predictor variable in bilingual acquisition studies (Unsworth 2013a).

In sum, in this study, input will be explored in terms of its cumulative amount since the beginning of instruction (length of instruction in years, and in curricular and extracurricular hours), amount of time spent in a naturalistic immersion situation involving native speaker contact, and current informal contact with the target language, the latter two involving linguistic richness as well. The research questions of the study are as follows:

1 What is the strength of the association between L2 oral performance with starting age, on the one hand, and with input, on the other, in learners with a long learning experience (more than 10 years)?
2 Which input measures (length of instruction in years, number of curricular and extracurricular lesson hours, amount of time spent in a naturalistic immersion situation abroad, current informal contact with the target language) are more strongly associated with long-term L2 oral performance?

On the basis of the previous study, the prediction is made with respect to the first research question that input measures will show a stronger correlation with L2 oral performance than starting age. However, with respect to the second question, the absence of research evidence precludes a prediction.

Participants

The participants in this study were 160 participants from two different universities in Spain (including some of the participants in the previous analysis), of which 127 were female and 33 male. They were undergraduate students, many of them majoring in English, with an intermediate to advanced level of English. Most of them were multilingual (4 per cent knew two languages, 35 per cent knew three, and 61 per cent knew four or more languages). This group had had at least 10 years of English language learning experience; the average length was 15.5 (*SD* 3.6), and the median 15.1. The mean starting age, defined as the beginning of exposure to English as an FL (preschool, primary school, or secondary school) was 7.67 years (*SD* 2.2) and the range 3–15.5.[4] Their average age at testing was 22.76 years (*SD* 3.9), and the median was

21 years; no participant older than 35 years was included to avoid confounding effects from their biological age (i.e. aging effects; see Birdsong 2006, 2014) and to guarantee continuity in length of exposure.

Procedure and instruments

Participants completed an extensive questionnaire concerning their English learning history including questions about age of onset of instruction and quantity and type of input they received in the different levels of education, both formally and informally (out-of-school exposure).[5] As seen above, the measures used in this study included measures of cumulative exposure: length of instruction in years, length of instruction in curricular and extracurricular hours; and length of exposure in hours during stays abroad (more than 2 weeks) in an English-speaking country. A fourth measure was current frequency of contact with the target language outside the classroom. This was a composite measure calculated from the responses to four questions for which respondents were asked to rate frequency in a scale from 0 (never) to 5 (daily): watching TV and films in English; writing emails, letters, etc. in English; reading extended texts in English; and other experiences of intensive exposure the participants currently had. Significant responses included a variety of practices such as conversations with native or expert speakers and use in the workplace (but practice with non-experts/pupils for those participants who were school teachers of English was not considered significant practice). Both hours of immersion abroad and current contact include contact with native speakers and thus are indicators of input quality as well.

Language data were collected by means of an extensive test battery; some of the tests were computer-administered collectively, whereas some others were administered individually. The latter included the film-retelling oral narrative that provided the measures of oral performance, which will be analysed in the present article. Participants were asked to watch a clip of the film *Modern Times* once and then a second time in which they were asked to retell the story in two parts, at a mid-point in the story and at the end. Task performance happened in the presence of a researcher who was instructed not to interact with the participants.

All narrative samples were transcribed by means of the CLAN mode of the CHILDES database. In the analysis, five measures of performance were used that tapped into different dimensions: structural and lexical complexity, accuracy, and fluency (CAF). Fluency was measured by means of a measure of speech rate, number of pruned syllables per minute (repetitions, self-repairs, and false starts were eliminated). Lexical diversity, a dimension of lexical complexity, was measured by means of the D index (Malvern *et al.* 2004), which is not affected by text length and measures lexical diversity through a process of curve-fitting.[6] Structural complexity was measured on the basis of analysis of speech (AS) units, which take into account the features of spoken language. An AS unit 'is a single speaker's utterance consisting of an independent clause,

or subclausal unit, together with any subordinate clause(s) associated with either' (Foster *et al*. 2000: 365). Two measures were used: clauses per AS unit, which is a measure of subordination, and words per AS unit, as a measure of overall syntactic complexity. Accuracy was measured by means of errors per 100 words, which is a measurement often used in psycholinguistic research that has the advantage of compensating for differences in text or speech length. Interrater measures were used on 10 per cent of the data with overall interrater reliability reaching 96 per cent.

Results

Table 1 presents descriptive statistics of the measures drawn from the questionnaire concerning starting age, age at testing, cumulative measures of input (number of years since beginning of instruction, number of hours of curricular and extracurricular lessons, number of hours in a social immersion context abroad), and current frequency of informal contact with the target language. Table 2 displays the descriptive statistics of the measures used in the analysis of the oral narratives.

Pearson correlations between starting age and measures of input with the language measures were run to compare the strength of association of these variables. As can be seen in Table 3, no significant correlations were found between starting age and any of the performance measures.[7] In contrast, input measures showed several significant associations.

Specifically, the measure of current informal contact shows the highest number of significant correlations with oral performance measures: speech rate, lexical diversity, and the two measures of syntactic complexity. Next,

Table 1: Descriptive statistics. Age and input measures

Measures	Mean	*SD*	*N*
Starting age	7.7	2.2	159
Age at testing	22.8	3.9	160
LoI (years)[a]	15.5	3.6	160
LoI (C + EC h)[b]	2988.1	836.6	160
LoE abroad (h)[c]	1513.9	1580.8	108
Current L2 contact[d]	3.7	1.2	155

[a]Length of instruction in years.
[b]Length of instruction in curricular and extracurricular hours.
[c]Length of exposure during stays abroad in hours.
[d]Current frequency of informal contact with the target language on a scale from 0 (never) to 5 (daily).

Table 2: Descriptive statistics. Oral performance measures

Measures	Mean	*SD*	*N*
Speech rate (pruned syllables per minute)	134.8	38.1	160
D index	52.1	10.6	160
Clauses × AS unit	2.1	0.6	160
Words × AS unit	10.5	1.9	160
Errors × 100 words	4.1	2.4	160

Table 3: Pearson r coefficients. Input measures and oral performance measures

	Speech rate ($n = 100$)	D index ($n = 101$)	Clauses × AS unit ($n = 101$)	Words × AS unit ($n = 97$)	Errors × 100 words ($n = 99$)
Starting age	—	—	—	—	—
LoI (years)[a]	—	—	.242**	.288**	—
LoI (C + EC h)[b]	.163^{+}	—	—	—	−.160^{+}
LoE abroad (h)[c]	.360***	.298**	—	—	−.289**
Current L2 contact[d]	.352***	.215*	.333***	.250**	—

*Significant at .05 level (2-tailed); **significant at .01 level (2-tailed); ***significant at .001 level (2-tailed); $^{+}$marginally significant (.052 and .057, respectively).
[a]Length of instruction in years.
[b]Length of instruction in curricular and extracurricular hours.
[c]Length of exposure during stays abroad in hours.
[d]Current frequency of informal contact with the target language on a scale from 0 (never) to 5 (daily).

the measure of number of hours in an English-speaking community abroad significantly correlates with the measures of oral fluency and of lexical diversity, as well as with the measure of accuracy, though it does not correlate with the two measures that indicate syntactic complexity. In contrast, the measure of number of years since beginning of instruction only correlates with the two measures of syntactic complexity. Finally, the measure number of hours of curricular and extracurricular lessons shows only marginal correlations.

To have a more complete comparison of the relative effects of the different input measures on oral performance, a set of standard multiple regression analyses were conducted. The five oral performance measures were

dependent variables, and the four input measures the predictive variables. All the assumptions for regression were met: no multicollinearity between predictive variables (they were all moderate to low); outliers on both predictive and dependent variables were detected and deleted from the dataset for each regression analysis.

The variable number of hours of curricular and extracurricular lessons, which had shown marginally significant correlations, did not emerge as a significant predictor of any of the CAF measures. The other three independent variables predicted the different dimensions of oral performance in different ways.

Fluency: speech rate

The results of the regression indicated that the input measures predicted 19 per cent of the variance ($R^2 = .19$, $F(4,95) = 5.69$, $p < .001$). The best predictors were hours spent abroad and current informal contact. It was found that hours spent abroad significantly predicted speech rate ($\beta = .249$, $t(95) = 2.47$, $p < .05$), as did current contact ($\beta = .253$, $t = 2.53$, $p < .05$). They uniquely accounted for 5.2 and 5.5 per cent, respectively, of the variance in the model.

Lexical diversity: D index

The results of the regression indicated that the input measures predicted 11 per cent of the variance ($R^2 = .11$, $F(4,96) = 2.84$, $p < .05$). The best predictor was hours spent abroad, which significantly predicted lexical diversity ($\beta = .25$, $t(96) = 2.33$, $p < .05$), and uniquely accounted for 5 per cent of the variance in the model.

Syntactic complexity: subordination measured by means of clauses per AS unit

The results of the regression indicated that the input measures predicted 18 per cent of the variance ($R^2 = .18$, $F(4,96) = 5.43$, $p < .01$). The best predictors were current informal contact and number of years since beginning of instruction. The former significantly predicted sentential syntactic complexity ($\beta = .37$, $t(96) = 3.72$, $p < .001$), as did number of years ($\beta = .22$, $t(96) = 2.35$, $p < .05$). They uniquely accounted for 11.8 and 4.7 per cent, respectively, of the variance in the model.

Overall syntactic complexity: words per AS unit

The results of the regression indicated that the input measures predicted 13 per cent of the variance ($R^2 = .13$, $F(4,92) = 3.42$, $p < .05$). The best predictors were number of years since beginning of instruction and current informal contact. The former significantly predicted overall syntactic complexity ($\beta = .26$,

$t(92) = 2.64$, $p < .001$), as did current contact ($\beta = .22$, $t = 2.19$, $p < .05$). They uniquely accounted for 6.5 and 5.6 per cent, respectively, of the variance in the model.

Accuracy: errors per 100 words

The results of the regression indicated that the input measures predicted 10 per cent of the variance ($R^2 = .10$, $F(4,94) = 2.62$, $p < .05$). There was only one variable contributing to the variance, hours spent abroad. It significantly predicted accuracy ($\beta = -.26$, $t = 2.42$, $p < .05$), accounting for 5.6 per cent of the variance in the model.

DISCUSSION

The first aim of this study was to explore the strength of the association of learners' starting age and L2 oral performance in the long term and to compare it with the strength of the association between input measures and L2 oral performance. A minimum period of 10 years of continuous instruction was taken as a realistic indicator of long-term achievement in an FL setting. The correlational analyses carried out with starting age, input measures, and oral performance measures confirmed the prediction that in the long term certain input characteristics are more tightly associated to learners' L2 oral performance than starting age. It was found that starting age did not correlate significantly with any of the measures of oral performance in an oral narrative task in which early starters could have used implicit knowledge (Ellis 2005) acquired in childhood. This supports previous results from different types of task that also showed a lack of association of starting age with proficiency measures (Al-Thubaiti 2010; Muñoz 2011).

A caveat needs to be mentioned because the amount of exposure received by these learners is still far smaller than the amount of exposure received by learners in a naturalistic situation after 10 years of residence, and it may be argued that younger starters may need a still longer period to catch up to older starters. However, the length of exposure included in this study (an average of >15 years) approximates the upper limit of students' period of instruction, and thus it marks the realistic end for research as well. In addition, the lack of any correlation seems to indicate that the older starters in the sample do not obtain higher scores than the younger starters, as they no longer have a cognitive maturation advantage when learners are all in their early twenties.

These results support the view of learning context as a mediating factor in L2 acquisition (Muñoz 2008). First of all, the young learners' slow pace of learning that has been consistently shown by research may be related to the scarcity of input in typical FL learning settings. As DeKeyser (2000) argued, young children are good at implicit learning, but implicit learning mechanisms need massive amounts of input. If the learning setting does not offer the amount of input required, implicit learning will not be facilitated. In contrast,

older children and adolescents will be better at explicit learning because of their superior cognitive maturity. This older learner advantage is favoured by typical school settings, which results in an even greater advantage of older over younger instructed learners. In the long term, younger starters will not show an advantage either because they will not have benefited from learning implicitly due to input limitations, whereas older starters will have lost their relative cognitive maturity advantage (Muñoz 2006b). This explains that starting age does not appear to be a determining factor of long-term oral performance of the learners in this study even in a task in which early starters could have used implicit knowledge acquired in childhood.

The second aim of the study was to explore which input measures were the best predictors of L2 oral performance in these learners. Significant correlations between input measures and oral performance measures were unsurprisingly modest, indicating that the contribution of other internal (e.g. aptitude and motivation) and external variables (e.g. teaching quality, teachers' oral proficiency) is needed to explain learners' variability in oral performance more completely. Nevertheless, regression analyses showed interesting results that set apart the different input measures used in this study. First, both the measure of current informal contact and of hours of immersion abroad emerged as better predictors than the measurements of length of instruction. This highlights the importance of contact with native speakers and exposure to input that is linguistically rich. The relatively high weight of the measure of current informal contact with the language also highlights the central role played by learners' orientations and engagement with the language at the time of the study (see Muñoz and Singleton 2011). As suggested by Moyer (2004: 140, 2014), variety of contact sources (formal and informal, personal and professional domains) and frequency of personal contact result in more opportunities to use the L2 and greater confidence and sense of self in the language, which ultimately lead to more practice opportunities and increased fluency in the language. In comparison, length of learning/instruction is a weaker predictor of L2 achievement.

Another interesting result is the different associations revealed between input measures and oral performance dimensions. First, number of hours of immersion in the L2 abroad is associated to higher fluency, lexical diversity, and accuracy. This is in alignment with findings of research that has investigated gains obtained in study abroad experiences, which have most frequently reported gains in fluency (e.g. Freed 1995; Juan-Garau and Pérez-Vidal 2007; Llanes and Muñoz 2009, 2013). Other studies have also shown gains in lexical diversity during a study abroad experience, for example, the study by Foster (2009), also using the D index. Although gains in accuracy have not always been reported, some studies have found significant decrease of errors too, particularly lexical errors (Llanes 2012). However, number of hours of immersion in the L2 abroad is not a significant predictor of syntactic complexity, which suggests that naturalistic immersion does not contribute to increasing learners' syntactic complexity as much as other types of learning

experience in these learners. Study abroad research has shown conflicting results (e.g. Collentine 2004; Llanes and Muñoz 2013) in relation to this performance dimension.

Secondly, the measure of current contact with the target language appears as a good predictor of fluency, as good as hours spent abroad, which concurs with the findings of Freed *et al*. (2004) showing that students in an immersion setting at home improved their oral fluency even more than students abroad because the former reported more hours of productive use of the L2. This measure also appears as a good predictor of syntactic complexity, in contrast to hours abroad.

The third measure, number of years of instruction, appears as a significant predictor of the two measures of syntactic complexity, the measure of subordination and the overall measure of complexity. This may be a reflection of the characteristics of formal instruction, with an emphasis on grammar over fluency or lexical diversity. Furthermore, the positive impact that instruction might have on syntactic complexity may also be inferred from the finding that differences in gains are comparatively smaller in this dimension when comparing students who have stayed abroad with those who have stayed in their home university (e.g. Llanes and Muñoz 2013).

These results support those of the previous study that used proficiency measures (Muñoz 2011), in which also the number of hours abroad and current L2 contact showed a stronger correlation with the scores in the global proficiency test than the number of years since beginning of instruction.

To finish, it is interesting to compare the role of this general measure of exposure in an FL learning setting and a naturalistic L2 setting. First, it appears that the impact of cumulative exposure may be inversely proportional to the amount of input in the environment: LOR is not as strongly associated with L2 achievement in a naturalistic setting (e.g. Johnson and Newport 1989) as in an FL setting (Muñoz 2008). This may suggest that when exposure to the target language is unlimited, once learners have reached a certain amount of immersion (e.g. 10 years), input ceases to be a determinant factor, but this is not the case when exposure is limited. However, it appears that even in situations of naturalistic exposure, the importance of LOR is dependent on the amount of contact with native speakers (or near-native speakers) and the opportunities for significant interaction with them, as shown by studies indicating that LOR is a less determining factor than high-quality input (Flege and Liu 2001; Jia and Fuse 2007; Flege 2009). This is in line with the findings of the present study indicating that hours of immersion in an English-speaking environment and current contact with English speakers at home are more deterministic factors for oral performance in the long term than number of years of FL learning. It may also be added that during those years learners may have been exposed to classroom input of varying quality, from input that is inconsistent and makes the acquisition task more difficult (Rothman and Guijarro-Fuentes 2010) to input that provides an adequate model for acquisition.

CONCLUSION AND LIMITATIONS

This study has explored the relationship between starting age, input, and learners' oral performance, thereby enriching our knowledge of the effects of age and input in language learning settings where input is limited. Specifically, the study contributes to filling the gap concerning oral performance in the line of research that investigates long-term effects of starting age, and to discerning the sources and types of input that may bear a stronger influence on the different oral performance dimensions. The results of the study have confirmed that cumulative exposure and, above all, input quality and contact with native speakers are more deterministic factors than mere starting age.

The study has a number of limitations. The fact that most participants were multilingual may be seen as a limitation with respect to the generalizability of the findings. Another characteristic of the sample is the limited variability in learners' starting age, given the type of sample needed. In this regard, it may be argued that significant correlations may be difficult to obtain because of the narrow range in starting age, but also the variability in number of years since beginning of instruction was small and, in this case, a few correlations reached significance. Further research that includes starting age, input, and other key factors such as aptitude and motivation is needed to complete our understanding of long-term achievement in FL learning.

The educational implications of these findings are manifold. First, although an early introduction of FL instruction may be well justified because of the additional time for learning that it provides and other educational motives (Johnstone 2007), unrealistic expectations should not be held with respect to attainment levels when learners do not have abundant and rich exposure to the FL. Secondly, because intensive exposure seems to be more effective than long periods of drip-feed instruction, the convenience of educational programmes that integrate periods of intensive exposure at home and/or abroad is supported by the findings in this study. Immersion programmes, abroad or at home that provide naturalistic or quasi-naturalistic exposure, will be especially beneficial at a young age because of the combination of greater possibilities for implicit learning, at which young learners have an advantage (DeKeyser 2000), of intensive input and of more plentiful opportunities for interaction with native-speaker peers (Muñoz and Llanes 2014).

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge research grants FFI2010-21478, 2014SGR1089 and ICREA Academia.

NOTES

1 An approximate calculation with four 1-h sessions per week gives 245 years.

2 Ojima *et al.* (2011: 198) describe ERPs as 'online electrical measures of neural

activities, obtained by timelocking scalp-recorded electroencephalograms to a certain event such as the presentation of linguistic stimuli.' Further, 'ERPs can visualize language processing as it is occurring, and provide online data of early linguistic processes well before any conscious decision can be made.'

3 Note that the term *cumulative* is not used in its conventional sense in Unsworth (2013a), where it entails proportions of the use of each language in bilingual speakers. In this article, however, the term *cumulative* is used in its conventional sense of exposure over time.

4 A limitation needs to be noted that seems unavoidable in this line of non-experimental research: because of restrictions on minimum length of exposure (10 years) and maximum age at testing (30 years), late learners are under-represented in the sample.

5 A drawback of this type of study is that it relies on participants' remembering the number of weekly hours of extracurricular lessons; in contrast, the number of curricular hours of English instruction per week/year has little variation in the public school system in Spain.

6 The D index is an improvement on measures that are affected by text length. It is held to be more informative than the type/token ratio because it represents how it varies over a range of token sizes for each speaker or writer. For the computation of D, the instrument *D_Tools* (Meara and Miralpeix 2008) was used.

7 In case cumulative measures of input were strongly associated to starting age, which could hide the effects of this factor, correlations between these variables were checked. Starting age was only moderately correlated with number of years ($r = -.162$; $p < .05$) not with any of the other input measures. As expected, partial correlations between number of years and the learners' oral performance measures controlling for starting age showed no significant changes.

REFERENCES

Abrahamsson, N. and **K. Hyltenstam.** 2008. 'The robustness of aptitude effects in near-native second language acquisition,' *Studies in Second Language Acquisition* 30/4: 481–509.

Al-Thubaiti, K. A. 2010. 'Age effects in a minimal input setting on the acquisition of English morpho-syntactic and semantic properties by L1 speakers of Arabic,' Unpublished doctoral dissertation, University of Essex.

Birdsong, D. 2006. 'Age and second language acquisition and processing: A selective overview,' *Language Learning* 56/1: 9–49.

Birdsong, D. 2014 (in this issue). 'Dominance and age in bilingualism,' *Applied Linguistics.* doi:101093/applin/amu031.

Birdsong, D. and **M. Molis.** 2001. 'On the evidence for maturational constraints on second-language acquisition,' *Journal of Memory and Language* 44: 235–49.

Burstall, C. 1975. 'Primary French in the balance,' *Foreign Language Annals* 10/3: 245–52.

Clark, E. V. 2003. 'Critical periods, time, and practice,' *University of Pennsylvania Working Papers in Linguistics* 9: 39–48.

Collentine, J. 2004. 'The effects of learning contexts on morphosyntactic and lexical development,' *Studies in Second Language Acquisition* 26: 227–48.

DeKeyser, R. M. 2000. 'The robustness of critical period effects in second language acquisition,' *Studies in Second Language Acquisition* 22: 499–533.

DeKeyser, R. M. 2013. 'Age effects in second language learning: Stepping stones toward better understanding,' *Language Learning* 63 (Suppl. 1), 52–67.

DeKeyser, R. M., I. Alfi-Shabtay, and **D. Ravid.** 2010. 'Cross-linguistic evidence for the nature of age effects in second language acquisition,' *Applied Psycholinguistics* 31: 413–38.

Ellis, R. 2005. 'Measuring implicit and explicit knowledge of a second language,' *Studies in Second Language Acquisition* 27: 141–72.

European Commission. 2012. *First European Survey on Language Competences.* Education and Training.

Flege, J. E. 2009. 'Give input a chance!' in Piske and Young-Scholten (eds).

Flege, J. E. and **S. Liu.** 2001. 'The effect of experience on adults' acquisition of a second language,' *Studies in Second Language Acquisition* 23: 527–52.

Flege, J. E., D. Birdsong, E. Bialystok, M. Mack, H. Sung, and **K. Tsukada.** 2005. 'Degree of foreign accent in English sentences produced by Korean children and adults,' *Journal of Phonetics* 33: 263–90.

Foster, P. 2009. 'Lexical diversity and native-like selection: The bonus of studying abroad' in B. Richards, M. Daller, D. Malvern, P. Meara, J. Milton, and J. Treffers-Daller (eds): *Vocabulary Studies in First and Second Language Acquisition*. Palgrave Macmillan.

Foster, P., A. Tonkyn, and **G. Wigglesworth.** 2000. 'Measuring spoken language,' *Applied Linguistics* 21/3: 354–75.

Freed, B. 1995. 'What makes us think that students who study abroad become fluent?' in B. Freed (ed.): *Second Language Acquisition in a Study Abroad Context*. John Benjamins.

Freed, B., N. Segalowitz, and **D. Dewey.** 2004. 'Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs,' *Studies in Second Language Acquisition* 26: 275–301.

García Mayo, M. P. and **M. L. García Lecumberri** (eds). 2003. *Age and the Acquisition of English as a Foreign Language: Theoretical Issues and Field Work*. Multilingual Matters.

Gathercole, V. 2007. 'Miami and North Wales, so far and yet so near: Constructivist account of morpho-syntactic development in bilingual children,' *International Journal of Bilingual Education and Bilingualism* 10/3: 224–47.

Gathercole, V. and **E. Hoff.** 2007. 'Input and the acquisition of language: Three questions' in E. Hoff and M. Shatz (eds): *The Handbook of Language Development*. Blackwell.

Gathercole, V. and **E. Thomas.** 2009. 'Bilingual first-language development: Dominant language takeover, threatened minority language take-up,' *Bilingualism: Language and Cognition* 12/2: 213–37.

Granena G. and **M. H. Long** (eds). 2013. *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment*. John Benjamins.

Hoff, E., C. Core, S. Place, R. Rumiche, M. Señor, and **M. Parra.** 2012. 'Dual language exposure and early bilingual development,' *Journal of Child Language* 39: 1–17.

Jia, G. and **A. Fuse.** 2007. 'Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences,' *Journal of Speech, Language, and Hearing Research* 50: 1280–99.

Jia, G., W. Strange, Y. H. Wu, J. Collado, and **Q. Guan.** 2006. 'Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure,' *The Journal of the Acoustical Society of America* 119: 1118–30.

Johnson, J. and **E. Newport.** 1989. 'Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language,' *Cognitive Psychology* 21: 60–99.

Johnstone, R. 2007. 'Nationally sponsored innovations at school in Scotland: Issues of evidence, generalisability and sustainability,' *Innovation in Language Learning and Teaching* 1/1: 111–18.

Juan-Garau, M. and **C. Pérez-Vidal.** 2007. 'The effect of context and contact on oral performance in students who go on a stay abroad,' *Vigo International Journal of Applied Linguistics* 4: 117–34.

Kalberer, U. 2007. 'Rate of L2 acquisition and the influence of instruction time on achievement,' M.Ed. thesis, University of Manchester.

Krashen, S., M. Long, and **R. Scarcella.** 1979. 'Age, rate and eventual attainment in second language acquisition,' *TESOL Quarterly* 9: 573–82.

Larson-Hall, J. 2008. 'Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation,' *Second Language Research* 24/1: 35–63.

Lenneberg, E. H. 1967. *Biological Foundations of Language*. Wiley.

Llanes, À. 2012. 'The impact of study abroad and age on second language accuracy development' in C. Muñoz (ed.).

Llanes, À. and **C. Muñoz.** 2009. 'A short stay abroad: Does it make a difference?,' *System* 37: 353–65.

Llanes, À. and **C. Muñoz.** 2013. 'Age effects in a study abroad context: children and adults studying abroad and at home,' *Language Learning* 63: 63–90.

Long, M. H. 2007. *Problems in SLA*. Lawrence Erlbaum.

Long, M. H. 2013. 'Some implications of research findings on sensitive periods in language learning for educational policy and practice' in Granena and Long (eds).

Malvern, D., B. Richards, N. Chipere, and **P. Durán.** 2004. *Lexical Diversity and Language Development: Quantification and Assessment*. Palgrave.

Meara, P. and **I. Miralpeix.** 2008. *D_Tools* (Version 2.0) [Computer software]. Lognostics.

Moyer, A. 2004. *Age, Accent and Experience in Second Language Acquisition. An Integrated Approach to Critical Period Inquiry*. Multilingual Matters.

Moyer, A. 2009. 'Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment' in Piske and Young-Scholten (eds).

Moyer, A. 2014 (in this issue). 'Exceptional outcomes in L2 phonology: the critical factors of learner engagement and self-regulation,' *Applied Linguistics.* doi:10.1093/applin/amu012.

Munro, M. and **V. Mann.** 2005. 'Age of immersion as a predictor of foreign accent,' *Applied Psycholinguistics* 26: 311–41.

Muñoz C. (ed.). 2006a. *Age and the Rate of Foreign Language Learning*. Multilingual Matters.

Muñoz, C. 2006b. 'The effects of age on foreign language learning: The BAF Project' in C. Muñoz (ed.).

Muñoz, C. 2008. 'Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning,' *Applied Linguistics* 29: 578–96.

Muñoz, C. 2011. 'Is input more significant than starting age in foreign language acquisition?,' *International Review of Applied Linguistics* 49: 113–33.

Muñoz C. (ed.). 2012a. *Intensive Exposure Experiences in Second Language Learning*. Bristol: Multilingual Matters.

Muñoz, C. 2012b. 'The significance of intensive exposure as a turning point in learners' histories' in C. Muñoz (ed.).

Muñoz, C. and **À. Llanes.** 2014. 'Study abroad and changes in degree of foreign accent in children and adults,' *The Modern Language Journal* 98: 432–49.

Muñoz, C. and **D. Singleton.** 2011. 'A critical review of age-related research on L2 ultimate attainment,' *Language Teaching* 44/1: 1–35.

Myles, F. and **R. Mitchell.** 2012. 'Learning French from ages 5, 7 and 11: An investigation into starting ages, rates and routes of learning amongst early foreign language learner,' ESRC End of Award Report RES-062-23-1545.

Ojima, S., H. Matsuba-Kurita, N. Nakamura, T. Hoshino, and **H. Hagiwara.** 2011. 'Age and amount of exposure to a foreign language during childhood: Behavioral and ERP data on the semantic comprehension of spoken English by Japanese children,' *Neuroscience Research* 70: 197–205.

Paradis, J. 2010. 'Bilingual children's acquisition of English verb morphology: Effects of language dominance, structure difficulty, and task-type,' *Language Learning* 60/3: 651–80.

Paradis, J. 2011. 'Individual differences in child English second language acquisition: Comparing child-internal and child-external factors,' *Linguistic Approaches to Bilingualism* 1: 213–37.

Penfield, W. and **L. Roberts.** 1959. *Speech and Brain Mechanisms*. Atheneum.

Piske T. and **M. Young-Scholten** (eds). 2009. *Input Matters in SLA*. Multilingual Matters.

Rothman, J. and **P. Guijarro-Fuentes.** 2010. 'Input quality matters: Some comments on input type and age-effects in adult SLA,' *Applied Linguistics* 31: 301–6.

Singleton, D. 1989. *Language Acquisition: The Age Factor*. Multilingual Matters.

Singleton, D. 2012. 'Ultimate attainment and the Critical Period Hypothesis: Some thorny issues' in M. Watorek, S. Benazzo, and M. Hickmann (eds): *Comparative Perspectives on Language Acquisition*. Multilingual Matters.

Snow, C. 1983. 'Age differences in second language acquisition: Research findings and folk psychology' in K. M. Bailey, M. H. Long, and

S. Peck (eds): *Second Language Acquisition Studies*. Newbury House Publishers.

Snow, C. and **M. Hoefnagel-Höhle.** 1978. 'The critical period for language acquisition: Evidence from second language learning,' *Child Development* 49: 1114–28.

Unsworth, S. 2013a. 'Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender,' *Bilingualism: Language and Cognition* 16/1: 86–110.

Unsworth, S. 2013b. 'Current issues in multilingual first language acquisition,' *Annual Review of Applied Linguistics* 33: 21–50.

Unsworth, S., E. Argyri, L. Cornips, A. Hulk, A. Sorace, and **I. Tsimpli.** In press. 'On the role of age of onset and input in early child bilingualism in Greek and Dutch,' *Applied Psycholinguistics*. DOI: 101017/S014271641200574.

Unsworth, S., K. de Bot, L. Persson, and **T. Prins.** 2012. 'Foreign Languages in Primary School Project,' Proceedings of the Foreign Languages in Primary Schools Projects: presentation results FLiPP-research, Amersfoort, 13 December 2012.

NOTES ON CONTRIBUTOR

Carmen Muñoz is a full-time Professor of Applied English Linguistics at the University of Barcelona, Spain. Her research interests include the effects of age and context on second-language acquisition, young learners in instructed settings, individual differences, and bilingual/multilingual education. Recent publications include the volume *Intensive Exposure Experiences in Second Language Learning* (Multilingual Matters, 2012) and *Age Effects in a Study Abroad Context: Children and Adults Studying Abroad and at Home* (Language Learning, 2013; co-authored). *Address for correspondence*: Carmen Muñoz, Department of English and German Philology, Universitat de Barcelona, Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain. <*munoz@ub.edu*>