



HAL
open science

Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation

Jenifer Larson-Hall

► **To cite this version:**

Jenifer Larson-Hall. Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*, 2008, 24 (1), pp.35-63. 10.1177/0267658307082981 . hal-00570738

HAL Id: hal-00570738

<https://hal.science/hal-00570738>

Submitted on 1 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation

Jenifer Larson-Hall *University of North Texas*

Received July 2006; revised December 2006; accepted January 2007

This study examined whether a younger starting age is advantageous in a situation of minimal exposure to an instructed foreign language (≤ 4 hours classroom contact per week). Previous theoretical and empirical studies indicated there should be no advantage for an earlier start. Japanese college students who started studying English between ages three and twelve ($n = 61$) were examined on a phonemic discrimination ($\text{ɹ}/\text{l}/\text{w}$) and grammaticality judgement task (GJT). After controlling for language aptitude and amount of input, statistical correlations were found between starting age and scores on the GJT ($r = -.38$) but not the phonemic task ($r = .03$). These earlier starters were also compared to peers who began study in junior high at age twelve or thirteen ($n = 139$) on the same measures. The earlier starters were found to score statistically higher on the phonemic but not morphosyntactic measure, and this remained true in an ANCOVA analysis where total amount of hours of study input were controlled for. A robust ANCOVA testing for differences at different levels of input found interesting interactions between group affiliation and amount of input. Language attitudes were also tested. The evidence shows there can be perceivable age effects for linguistic measures even in a situation of minimal exposure to a foreign language, but these may not emerge until a substantial amount of input has been gained.

Keywords: second language learning, age effects in L2, input effects on L2, L2 English phonology, L2 English morphosyntax

I Introduction

The question of what age is the optimal age to start learning a foreign language has vast practical and economic consequences on the resources of individuals and schools. However, as second language acquisition researchers we are far from the goal of clearly understanding the role that age plays when the amount of language input is very small. The present study examines the question of whether there is any advantage to starting study of a foreign language at an early age by looking at the performance of present-day Japanese college students on one phonological and one morphosyntactic measure and considering their language learning histories. One group of students began studying English at a median age of nine, while the other group began their study when they entered junior high school at age 12 or 13. The study is limited to learners who received what I call 'minimal input', defined as no more than four hours of instruction per week. Certainly, in such a situation learners will not acquire a nativelike proficiency in any area, but the issue is whether such study can ultimately advantage the language learners because of a younger starting age.

Lay people may assume that the robust finding of 'younger is better' applies to second language acquisition in all situations. However, all the critical period studies which show that 'younger is better' have been conducted with immigrant populations, in which child learners were immersed in the target language environment and presumably spent a large percentage of their waking hours being exposed to the second language in school (Oyama, 1976; Johnson and Newport, 1989; Flege *et al.*, 1999; DeKeyser, 2000). The 'younger is better' phenomenon has no guarantee of applying in situations of only minimal input; thus, it is imperative that we look at the long-run consequences of minimal input before making any conclusions.

Some researchers have argued that age effects would not be expected to appear in situations of minimal input because critical period effects are only seen in naturalistic or immersion classroom settings (Patkowski, 1994; DeKeyser, 2000; Lightbown, 2000). Theoretically, it could be argued that if children and adults learn through fundamentally different mechanisms (Bley-Vroman, 1988), and that in particular children learn primarily in an implicit way (DeKeyser, 2003), then small amounts of

input will not be enough to trigger the formation of a morphological, syntactic or phonological system.

Notice that this theoretical position might imply that younger starters would perform even worse than later starters after the same amount of input. Muñoz (2001) posits that earlier starters cannot gain the positive effects of an early start if there is insufficient input for the kind of implicit learning that is done by children, while they are not cognitively advanced enough yet to benefit from explicit methods of instruction.

In order to examine the long-term effects of an earlier start, but in a minimal input situation, test results from 200 learners of English in Japan were obtained. These learners experienced at least 6 years of English study and were all at least 18 at the time of testing. The participants may not have reached their point of ultimate attainment in English, especially as many were continuing to study English at the college level, but this time period should be long enough for advantages, if there are any, to emerge.

In Japan, many parents spend large amounts of money to send their children to private language schools. Koike and Tanaka (1995) assert that Japan is one of the largest markets of foreign language teaching in the world, with an estimated 30 billion dollars spent on English language teaching every year. English language learning is viewed as important enough to future success that educators in Japan have recently decided to introduce compulsory English language instruction in the elementary schools (Butler, 2005). The answer to the question of whether early age is advantageous for these Japanese learners in a minimal input situation will be of interest to linguists as well as educators and administrators throughout many countries.

II Previous research on age advantages with minimal input

Several recent articles – using research done on two bilingual communities of third language learners of English in Spain (in the Basque Country and Catalonia) – find absolutely no linguistic benefits for younger starters (ages 4 or 8) vs. later starters (age 11) in a foreign language environment (García Mayo and García Lecumberri, 2003). Studies in the Basque Country (Cenoz, 2003; García Lecumberri and Gallardo, 2003; García Mayo, 2003; Lasagabaster and Doiz, 2003) examined students who had begun receiving 2–3 hours of English

instruction per week at ages 4, 8 and 11 (not all studies examined the four-year-old starters). Students were tested after 600 hours of school exposure, at testing ages of 9, 13 and 16. Similarly, the study on bilinguals studying English in Catalonia (Muñoz, 2003) looked at students who began at either age 8 or 11, and were tested once after 200 hours (at ages 11 and 13) and again after 416 hours (at ages 13 and 15). These participants thus seem to be quite comparable to those in the present study, in the sense that English was studied as a foreign language and input was minimal. The authors of the articles tested a variety of modalities and skill areas, but found no advantage for earlier starters in grammaticality judgements (García Mayo, 2003), degree of foreign accent or discrimination tasks (García Lecumberri and Gallardo, 2003), production and perception in oral testing (Cenoz, 2003; Muñoz, 2003) or composition writing (Cenoz, 2003). In fact, in almost every reported comparison it was found that earlier starters lagged behind later starters at least numerically and often statistically. In no cases were earlier starters statistically better than later starters. The only advantage found for earlier starters was in attitudes and motivation: Cenoz (2003) found statistically higher motivation levels among the youngest starters (4 years old) after 600 hours of instruction. However, in a different article on learners in Catalonia, Muñoz and Tragant (2001) found no difference in attitudes between 3rd grade (8–9 years) and 6th grade (11–12 years) starters.

The main problem with accepting the conclusion that early age gives little advantage for situations of minimal input using the studies in the García Mayo and García Lecumberri (2003) collection is that learners were tested after the same amount of exposure, meaning that the earlier starters were younger in chronological age than the later starters when they took the tests. There is evidence in psychology, specifically with IQ scores, that older learners will be better at test-taking than those who are cognitively more immature (Bernstein *et al.*, 1988) and, indeed, Cenoz (2003) notes that cognitive maturity may be a possible explanation for the findings of higher results overall for older starters. Thus, if the effect for an earlier start is small, it may be hidden by the cognitive advantages that the older starters hold on the tests. One might surmise that this explanation should then only apply on tests where explicit knowledge or test-taking is an issue (such as a grammaticality judgement test), and would not be expected to apply in situations of spontaneous production.

However, language learning strategies of memorizing vocabulary and conscious imitation of pronunciation models have been shown to be used more often by older than younger children (Victori and Tragant, 2003), and thus cognitive skills could also have an influence on other types of tests such as spontaneous production.

Older studies, conducted in the 1970s, also concluded that there was little linguistic advantage to beginning study of a second language earlier in a minimal input situation. Oller and Nagato (1974) examined Japanese students who began studying English in elementary school (1–2 hours a week) and compared them to peers who began English study in junior high only. Using a cloze test, the two groups were tested in 7th, 9th and 11th grades. The authors found statistical differences between the two groups in the lower two grades but not in 11th grade, and they concluded that the younger starters' advantages had disappeared by that point. The Oller and Nagato study thus avoids the problem of testing participants at different ages, but I would argue that a possible reason that no differences were found in 11th grade was statistical. The tests with 7th and 9th graders had approximately 50 students in each group, but the test in 11th grade had only 25 students who had started early. If effect sizes are small, statistical results can disappear in a smaller sample size (Tversky and Kahneman, 1971). For example, the power to find a medium effect size on a *t*-test with 50 participants is .7 (meaning there is a 70% chance of finding a difference if one exists), but the power to find the same effect size with only 25 participants is .4 (that is, chances are only 40% that you will find a difference even if one exists). The Oller and Nagato study has also been criticized because earlier starters were in the same classes with later starters, which means that if the earlier starters were more advanced in some areas, they had to 'wait' for the other learners to catch up. While this is a legitimate criticism of the study, it is not something that the present study avoids since, to my knowledge, regular Japanese schools do not distinguish levels of students when English classes begin in junior high. Thus, earlier starters would be put into the same classes as beginning learners.

Another well-known older study of British elementary students who learned French (by Burstall, 1977) had very large numbers of participants, but found that those who began learning French at age 8 showed no advantages over those who began at age 11 when both groups were

tested at age 16, except on a listening comprehension test. Even on the comprehension test, effect sizes were small (Burstall says the difference was 2–4 points on a 28-item test). However, students who started French earlier did have more positive attitudes towards French. Tests given in this study were quite holistic (at least for advanced students), and included reading, writing, speaking and listening comprehension tests. This study has the same problem as Oller and Nagato in that, for the most part, the earlier starters were put into the same classrooms as older students. However, Buckby (1976) shows that when students were separated into classes of earlier and later starters, at age 16 the earlier starters performed statistically better than older starters on listening and reading tests, and performed equally well on speaking and writing. Thus there did seem to be some advantages found for an earlier start but, overall, Burstall (1977) was pessimistic about the utility of a younger start for foreign languages. Although this is an important study, it is hard to compare its results with the present study, as the tests used were fairly different.

III A critical period for second language acquisition?

Extensive testing has established that ‘younger is better’ when it comes to both first and second language acquisition (for recent reviews, see Hyltenstam and Abrahamsson, 2003; DeKeyser and Larson-Hall, 2005). Large-scale studies such as Flege *et al.* (1999) have shown that age affects different parts of language differently, with phonology being affected at a much younger age than basic morphosyntax. Some authors have concluded there is a critical period for some aspects of lexical development as well (Mayberry and Eichen, 1991; Lederberg and Spencer, 2005). Even though the age phenomenon is robust, there is still considerable debate over the ultimate causes of this age-related phenomenon, with some claiming it is maturationally conditioned (Hyltenstam and Abrahamsson, 2003; DeKeyser and Larson-Hall, 2005), others arguing for more socio-psychological factors such as quantity and quality of input, motivation and affective variables (Jia *et al.*, 2002; Moyer, 2004), and some claiming that it is due to general cognitive declines that continue over the lifespan (Bialystok and Miller, 1999; Birdsong, 2005).

There are no definitive ages that have been pinpointed for possible critical period endpoints, but the Flege *et al.* (1999) study found that

even Koreans who had arrived in the USA as young as age 4 were, as a group, judged to pronounce sentences with more of an accent than a group of monolingual English speakers. From the scatterplot of foreign accent ratings given in that article (1999: 85), it can be noted that no Korean immigrant who arrived after the age of 12 scored in the range of native speakers. On the other hand, for a grammaticality judgement test (GJT) similar to the one used in the present study, Korean immigrants who arrived up to the age of 5½ were not significantly different from native speakers. For the scatterplot of scores for the GJT test (1999: 85), it is not until after age 18 that participants are not able to score within the range of native speakers.

There is really no reason to *a priori* privilege any particular age when doing a test of age effects, and the best methodology would be to use correlational analysis to look for trends instead of grouping participants and ‘throw[ing] away information’ (Maxwell and Delaney, 2004: 400). Such a methodology will be employed in this article for the group of younger starters, but because there is essentially no variation in age for the participants who began their study of English when they entered junior high, correlational measures will not be very informative. Therefore, I will also examine whether the younger starters differ from the older starters as a group.

Similarly to critical period studies, the present study compares long-term outcomes of younger vs. older starters. However, it is not a critical period study, as it does not compare the ultimate attainment of the learners and does not compare participants across a range of children to adults. It also differs from critical period studies because the context is that of English as a foreign language. However, critical period studies such as Johnson and Newport (1989) and DeKeyser (2000) provide examples of a methodology that could be used to compare the morphosyntactic abilities of large numbers of participants: the grammaticality judgement test (GJT).¹ Although such tests have been controversial (Sorace, 1996; Chaudron, 2003), researchers have experimentally found them to be reliable (Gass, 1994) and valid (Leow, 1996; Mandell, 1999), and many researchers continue to use them. The GJT test used in this study examines basic English

¹ Johnson and Newport (1989) tested 46 Chinese and Korean users of English and found a correlation of $r = -.77$ between age and GJT scores for all learners. DeKeyser (2000) tested 57 Hungarian users of English and found a correlation of $r = -.63$ between age and GJT scores across his entire group.

morphosyntax such as article usage, third-person singular marking and gender in pronouns. In their diagnostic role uncovering problems with basic morphosyntax that native speakers find essentially uncontroversial, this GJT is a much more direct and economical measure of grammatical intuition than spontaneous speaking or writing would be.

In order to obtain one specific measure of phonological ability, a receptive phonemic discrimination task was used. This listening test involved distinguishing whether words began with an initial *ɹ*/*l*/*w*. This phonemic discrimination task was previously used by Yamada *et al.* (1996) and found to be difficult for Japanese learners.² Thus, although such a task certainly does not take a measure of phonological ability as a whole, it was thought that it would be effective in separating learners on a difficult point of phonological ability.

Another measure that was used was a language learning aptitude test. Individual aptitude for learning a language has been shown to correlate with scores on linguistic tests for adults (Dörnyei and Skehan, 2003), and aptitude is thus an important factor that should probably be taken into account in all language acquisition studies that try to separate factors responsible for variation in scores. Oller and Nagato (1974) used an intelligence test in their study of younger vs. older starters and found a statistical correlation between IQ and cloze test scores ($r = .40$).

IV Research questions

The main issue investigated is whether an earlier starting age for studying English results in any benefits for Japanese learners when the learning context is that of receiving minimal input in a foreign language environment. Previous research and theory would indicate that there would be no advantage, and perhaps even a disadvantage, for a younger starting age in the minimal input situation, except for attitude towards language learning.

My first research question asks whether there is any direct relationship between the age that students began studying English and their scores on one phonological task and one morphosyntactic task. The answer to this question involves a correlational analysis among the earlier starters and can directly address the issue of age. In order to minimize the effects of

² Yamada *et al.* (1996) was designed to test the effect of word familiarity on discrimination abilities, since some of the words in their corpus were real and others were nonsense words. The authors found word familiarity correlated highly with perceptual ability ($r = .55$).

language learning aptitude as a possible confound, aptitude scores are factored out. As a further step in the correlational model, because there is a possible confound between age and amount of input – in that those who began learning at a younger age could have a larger amount of classroom input than those who began later – a partial correlation controlling for amount of input is also conducted.

My second research question is whether there are group differences between participants who began studying English before formal study in junior high school and those who did not. This result cannot speak directly to the issue of whether age itself confers any advantages, but does speak to the practical question of whether, as a group, any advantage can be found for Japanese students studying language before it becomes mandatory in junior high.

The third question of interest is whether there is any difference in attitude towards language learning in general or English in particular between the two groups of Japanese students. Given previous research, we would expect earlier starters to hold a more positive attitude towards learning language.

V Method

1 Participants

Over 200 native Japanese speakers who were college students at a national university in Japan took part in this study. The students tested here represent an elite group and can be considered among Japan's brightest students. Japan has seven original national universities, and getting into one of these universities is very difficult (Hughes, 1999).

The participants were students in 12 different classes and thus represent a sample of convenience, not a random sample. They were not paid for their participation, and they performed the tests and filled out the questionnaires during class time. No native speaker control group was included in this research design, as the question was not how the Japanese learners of English would perform relative to native speakers of English, but rather, whether an early start would provide any advantage over those with a later start.

Table 1, which summarizes some characteristics of the participants, can be found in the instruments section. Of the 200 participants, 102 were male and 98 female. They ranged from ages 18 to 21, with the average

age at testing overall of 19.4 years. Among the 61 students who reported studying English before junior high school ('earlier starters'), 20 were male and 41 were female. The majority of earlier learners (59%) began their study at age nine or later and studied English for 3 years or less. For those participants who began English study in junior high ('later starters'), their age of first exposure to English was 12 or 13, and they studied English for six years in school, as did the earlier learners. Fifty percent of all learners additionally studied English outside of school after they started junior high (66% of the earlier starters and 40% of the later starters). No information was gathered about socioeconomic status, although Burstall (1977) found this to correlate strongly with scores. In general, Japanese society is less economically stratified than western countries such as the USA, and Verba *et al.* (1987) reports there is only a weak correlation between educational level and income in Japan.

Although it is not possible to say with certainty what method was used to teach all of these Japanese learners of English, it is likely that English teaching in the state-funded schools focused mainly on vocabulary, non-integrated grammatical knowledge and translation exercises (Hughes, 1999; Morrow, 1987). The major reason for this is that, although the Japanese Ministry of Education states that it supports the goal of communicative language abilities, university entrance exams are comprised mainly of multiple choice items testing grammatical knowledge, translations and reading passages (Brown, 1995). Koike and Tanaka (1995) assert that teaching will be tailored towards doing those things that help students pass these difficult tests, and not towards achieving spoken fluency. Entrance exam test washback may not extend as far as conversation or cram schools for eight-year-olds, so it is difficult to know what kinds of teaching methodologies were employed for the earlier starters in this article, although practical considerations of age probably mean that teaching contained less explicit grammatical explanation.

2 *Instruments*

Students took an oral grammaticality judgement test (GJT), a phonemic discrimination test and an aptitude test.

a Grammaticality judgement test: The GJT used here was a version of Johnson and Newport's (1989) test of basic English morphosyntax,

adapted and used by DeKeyser (2000). The reliability – Kuder–Richardson formula 20 (KR-20) internal consistency estimate – of DeKeyser’s test administration was high, at over 90%. DeKeyser’s version was used without further change because most of the changes made for Hungarian learners are also challenging for Japanese learners. For example, the absence of determiners in Japanese make these sections difficult, while object pro-drop could also transfer from Japanese. This instrument was designed to test judgements on only basic features of morphosyntax in English that are unproblematic for native speakers. The 11 areas that DeKeyser (2000) lists in the GJT are:

- past tense (example: *‘Last night the old lady die in her sleep’)
- plural (*‘Teachers often give useful advices to their students’)
- third-person singular (*‘Our neighbor will goes to Europe next year’)
- present progressive (*‘Tom working in his office right now’)
- determiners (*‘Mrs. Johnson went to library yesterday’)
- pronominalization (pro-drop: *‘Mike wrote the letter but didn’t send’; gender errors: *‘John knew but she did not tell’)
- particle movement (*‘Kevin called Nancy for a date up’)
- subcategorization (depends on these individual verbs: say, laugh, say/tell, learn/study, want, hope, allow/let, enjoy; example: *‘I want you will go to the store now’)
- yes–no questions (*‘Can ride the little girl a bicycle?’)
- wh-questions (*‘Who you meet at the park every day?’)
- word order (*‘The student eats quickly his meals’)

This test was therefore meant to ascertain whether early exposure would prove an advantage for basic categories of morphosyntax. Although there is no guarantee that earlier starters learned these morphosyntactic features in their early classes, since these areas are very basic it seems likely that if they learned anything at all they would have been using these features of the grammar.

b Phonemic discrimination task: For the receptive phonological measure, Yamada *et al.*’s (1996) phonemic discrimination test was used. The test asked students to identify whether each of 96 items began with /r/l/w. Participants heard a real word like ‘ring’ or ‘wing’, or a nonsense word

like 'ling' (with the exception of the set royal/loyal/woyal, word groups consisted of only one syllable). Listeners heard each word only once, and there were 2 seconds between words. Between each group of 10 words there was an 8 second pause. Participants saw the sequence 'r l w' on their answer sheet and were told to circle the letter of the first sound of the word they heard. Yamada *et al.* found 99.8% accuracy on the test among 10 native speakers of American English, but only 66.9% accuracy among 50 Japanese speakers. A sample of some of the words can be found in Appendix 1. I do not have any information on how much attention was paid to pronunciation in classes for earlier starters except to note that among the early starters, 56% report that none of their teachers were native speakers, while 44% had at least one native speaker teacher.

c Aptitude test: The aptitude test was a Japanese version of the Modern Languages Aptitude Test (Carroll and Sapon, 1959) normed on 160 Japanese college students by Sasaki (1996). Only Parts 3 and 4, Language Analysis and Sound-Symbol Association, were used. The Language Analysis section consisted of 15 four-way multiple choice items, while the Sound-Symbol Association test had 22 four-way multiple choice items (18 written in katakana and four in Roman script). KR-20 estimates of reliability for Sasaki's administration were .80 for Language Analysis and .63 for Sound-Symbol Association.

d Background questionnaire: All students filled out a six-page questionnaire about their early experience with English, hours studying English in junior high and high school both in school and doing homework, attendance at language schools after age 12, self-ratings in speaking and reading, use of English, and attitude towards language (the questionnaire is available upon request). The self-ratings were measured on a 1–10 scale as shown below:

How well did you feel you could speak English on a conversational level before you entered this university? Place a **X** on the scale.

1.....	2.....	3.....	4.....	5.....	6.....	7.....	8.....	9.....	10
Can't	Can say	Can say	Can say	Can have	Can have	Don't have	Don't have	Like a	Like a
speak	words	simple	simple	a conversation	a conversation	to think	to think	native	native
at all	only	sentences	sentences			about	about	speaking	speaking
						speaking	speaking		

How well did you feel you could read English (in general, not in any specific field) before you entered this University? Put a **X** on the scale.

1.....	2.....	3.....	4.....	5.....	6.....	7.....	8.....	9.....	10
Can't	Can only	Can understand	Can read	Can read	Can read	Like a			
read	read some	simple sentences	paragraphs	easily	native				
at all	words					and naturally			

A rough probe of attitude was measured through two questions on the questionnaire. The questions were rated on a 1–10 scale where 1 = hate it, 5 = don't care, and 10 = love it. The questions were phrased: 'How much do you like studying languages?' and 'How much do you like studying English?' (these questions are similar to those in Gardner's AMTB for 'Interest in foreign languages' and 'Attitudes toward learning [a specific language]'; see appendix of Gardner, 1985). The use of English was based on the measure used in Flege *et al.* (1999) and asked about average use of English at home, school, part-time work and social situations on a 1–7 scale, where 1 = never and 7 = frequently.

For calculation of hours of exposure to the second language (L2), student reports of how many hours they spent per week studying English and also doing homework in both junior high and high school, and any time they studied English in formal classes outside school (such as in 'cram school') were noted. These numbers were multiplied by 44, for the number of weeks that students are in school. Because the students were recalling events from the previous 6–8 years, there will undoubtedly be some error in this measurement. However, student reports generally lined up with the Japanese ministry guidelines for how much English study is mandatory in junior high (three hours per week) and high school (five hours per week). About half of all participants (96 out of 200) reported studying English outside school in junior high and high school as well, and these hours were calculated into their total study time. It is also important to acknowledge that hours of instruction may not equal input. Recognizing the possible flaws of this variable it nevertheless seemed preferable to ask for self-report data in this case than to totally ignore possible differences in time spent exposed to English input. Measures for all participants can be furnished upon request.

Table 1 Summary characteristics of participants with mean and standard deviation in parentheses

	Earlier (<i>n</i> = 61)	Later (<i>n</i> = 139)
Years of study	3.0 (2.1) + 6*	6*
Amount of input (hours)	1923 (719)	1764 (727)
Age at testing	19.4 (0.9)	19.3 (0.8)
Age began formal instruction	8.3 (2.5)	12.5**
Self-rating (speaking; 10 = nativelike)	5.2 (1.1)	4.8 (1.1)
Self-rating (reading; 10 = nativelike)	6.9 (1.1)	6.6 (1.2)
Use of English (4–28 points, higher = more use)	7.1 (2.7)	7.0 (2.8)
Attitude toward studying language (10 = love it)	7.0 (1.9)	6.4 (1.9)
Attitude toward studying English (10 = love it)	7.1 (1.8)	6.5 (2.0)

* This number was not measured, but assumed as all students would have had to have this amount of study of English in the six years of junior high and high school.

** This is an estimated number as I did not ask those who began study in junior high what age they were when they entered junior high.

3 Procedure

The tests and questionnaires were given to students in their classrooms. The order of the three tests (the GJT, phonemic discrimination and aptitude) were varied in each class. Testing took 90 minutes: the GJT took 66 minutes, the aptitude test took 17 minutes and the phonemic discrimination test took 6 minutes. There were three versions of the phonemic discrimination test which were given in an ordered fashion (1–2–3) to different classes to avoid order effects.

VI Results

The main issue investigated is whether an earlier starting age for studying English results in any benefits for Japanese learners when the learning context is that of receiving minimal input in a foreign language environment. Previous research and theory would indicate that there would be no advantage, and perhaps even a disadvantage, for a younger starting age in the minimal input situation, except for attitude towards language learning.

1 Correlational analysis among the earlier starters

The first research question asks whether there is any direct relationship between the age that students began studying English and their scores on a phonological task and morphosyntactic task. The answer to this

question involves a correlational analysis among the earlier starters between age and test scores.

Before addressing the question of the relationship between age and test scores, however, I examined whether there were any correlations between language aptitude scores and test scores among the earlier starters. If there were a statistical correlation, we would want to factor aptitude out by doing a partial correlation, as Oller and Nagato (1974) did when they found a correlation between IQ scores and cloze test scores in their experiment. The correlation between aptitude scores and phonemic discrimination scores among earlier starters is $r = .37$, $n = 58$, $p < .05$, power = .84, and the correlation between aptitude scores and GJT scores is $r = .32$, $n = 59$, $p < .05$, power = .71.³ Because there is a statistical correlation between aptitude and both test scores, any further correlations will be partial correlations controlling for language learning aptitude.

Another factor to control for is amount of input. It may be the case that earlier starters just have a larger cumulative amount of input, and thus their advantage may have nothing to do with age *per se*. Figures 1 and 2 show scatterplots of the amount of input and scores on both tests. The lines superimposed on the graphs are smoother (Loess) lines and follow the trend of the data instead of fitting a straight line to the data like a regression line does. The scatterplots show that the data do not follow a straight line, and thus there may be no statistical correlation between input and test scores for the earlier learners: In fact, the correlation is not statistical for the phonemic discrimination task among the earlier starters ($r = -.07$, $n = 61$, $p = .57$, power = .08), but it is statistical for the GJT ($r = .27$, $n = 61$, $p = .035$, power = .56). However, it may still be worthwhile to examine how correlations change when amount of input is partialled out.

The correlation between scores on the phonemic discrimination task and starting age, controlling for aptitude, was not statistical ($r = .05$, $df = 58$, $p = .72$, $R^2 = .00$, power = .06). Controlling additionally for input, the situation does not change substantially and the correlation is still not statistical ($r = .03$, $df = 57$, $p = .80$, $R^2 = .00$). On the other hand, the

³ This is the value with outliers removed by using a robust correlation (Wilcox, 2001; Maronna *et al.*, 2006)

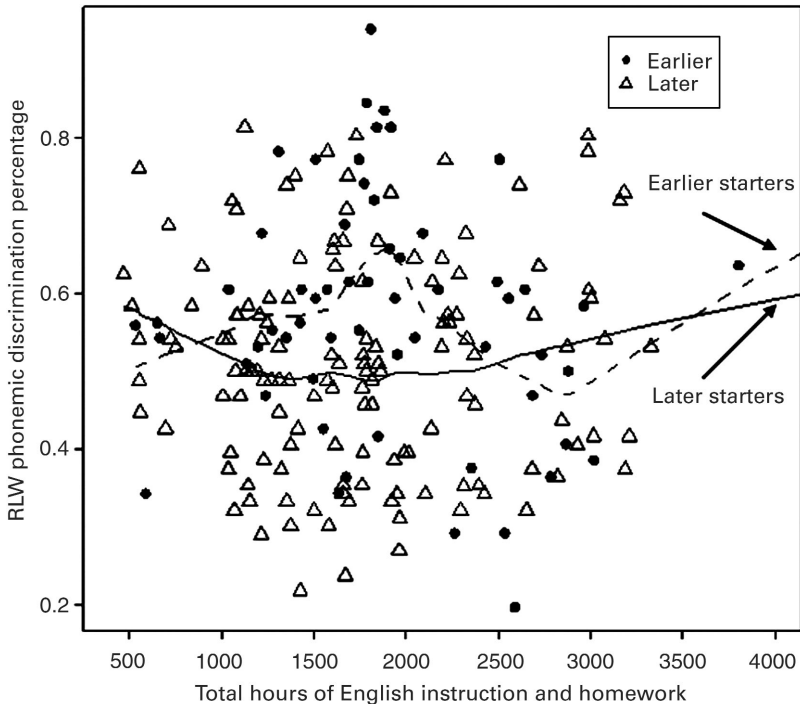


Figure 1 Scatterplot showing the relationship between amount of input and scores on the phonemic discrimination task for earlier and later starters

partial correlation (controlling for aptitude) between the grammaticality judgement test and starting age is statistical ($r = -.42$, $df = 58$, $p < .001$, $R^2 = .18$, power = .93). The negative correlation means that scores declined as the starting age increased. Controlling additionally for amount of input, there is only a slight decline in the strength of r ($r = -.38$, $df = 57$, $p = .003$, $R^2 = .14$).

2 *Examining group differences*

My second research question was whether a group of students who began their English study at a younger starting age would hold any advantage over those who only began English study in junior high. This is rather a different question than the question of the relationship of age to test scores, as was examined through the correlational analysis. Here, because the cut-off age of the groups is arbitrary, the question is not so

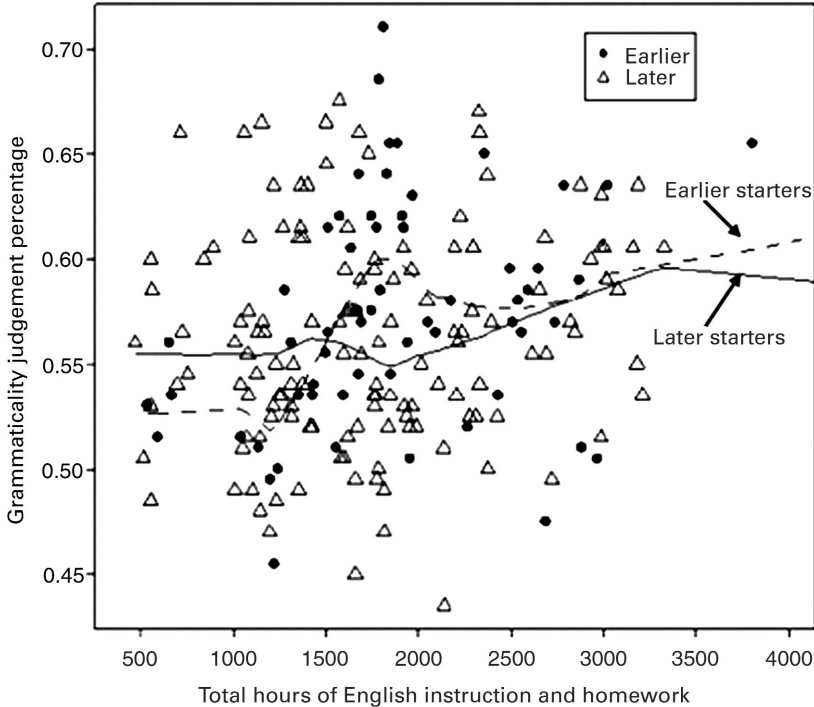


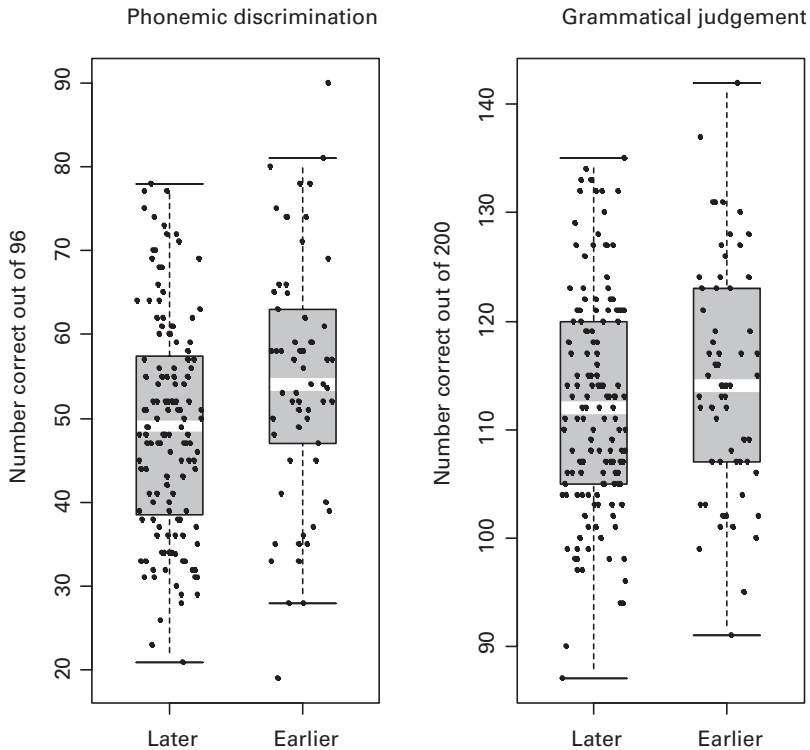
Figure 2 Scatterplot showing the relationship between amount of input and scores on the GJT for earlier and later starters

much whether age itself confers any benefits, but whether any benefits accrue to the group of students who were motivated enough (or whose parents were motivated enough!) to want to study English before it became mandatory in school. This still speaks to the practical question of whether any advantage can be found for Japanese students who begin their study of English before junior high with minimal input.

In comparing all of the earlier starters to all of the later starters, a series of three independent samples *t*-tests was conducted (a Bonferroni adjustment was applied and an alpha level of $.05/3 = .017$ was set). Mean scores and standard deviations on the GJT, phoneme discrimination test and aptitude test are given in Table 2. Skewness and kurtosis levels of the tests were within the normal range. A side-by-side graphic comparison of the groups can be seen in Figure 3. Results showed no statistical difference between groups for the aptitude test ($t_{120.4} = -.45$, $p = .65$, $\eta^2 = 0$; mean difference = $-.29 \pm 1.28$, power = .13). Although

Table 2 Results on 3 tests by overall group and earlier learners vs. later learners (standard deviation in parentheses)

	GJT (200 points)		R/L/W (96 points)		Aptitude (37 points)	
	Earlier	Later	Earlier	Later	Earlier	Later
Mean	114.7 (10.7)	112.3 (10.1)	54.7 (14.7)	49.3 (13.2)	31.6 (4.2)	31.2 (4.4)
Range	91–142	87–135	19–90	21–78	19–37	19–37

**Figure 3** Scores of Japanese learners of English compared on two measures; actual data points are overlaid on the boxplot and median lines are in white

we saw in the correlational analysis that there was a correlation between aptitude and test scores, the *t*-test shows that the difference in aptitude scores across groups is not statistical and thus should not affect group comparisons.

In looking at group differences on the two linguistic tests, there was no difference between groups on the GJT ($t_{109.2} = -1.51, p = .13, \eta^2 = .01$;

mean difference = -2.43 ± 3.19 , power = .82). There was, however, a statistical difference between the groups for the phonemic discrimination task ($t_{104,0} = -2.47$, $p = .015$, $\eta^2 = .03$; mean difference = -5.40 ± 5.40 , power = 1.00). The magnitude of the difference in the means (the η^2 value) for the phonemic discrimination test can be classified as small, according to Cohen (1988), and explains about 3% of the variance in the scores.

Again, as the earlier learners had spent more time involved in study and might have received more input, it is informative to examine group differences while controlling for total hours of input. In order to see whether there would be a difference between the earlier and later starters with equal amounts of input, a one-way ANCOVA with test scores (GJT or phonemic discrimination) as the dependent variable, status as earlier or later starter as independent variable, and total hours of input as the control variable was used to statistically match learners in both groups with equal amounts of input. An ANCOVA controls for the variability in scores due to the total hours studied, and is preferable to a matched pairs sample that may match learners unequally (Maxwell and Delaney, 2004). ANCOVA analysis confirmed the results of the t -tests, in that there was no effect for group (earlier vs. later starters) on the GJT ($F_{1, 197} = 1.69$, $p = .20$, partial $\eta^2 = .01$, power = .25), while there was an effect for group on the phonemic discrimination test ($F_{1, 197} = 6.55$, $p = .01$, partial $\eta^2 = .03$, power = .72). The covariate of total hours of study was statistical for the GJT ($F_{1, 197} = 6.24$, $p = .013$, partial $\eta^2 = .03$, power = .70), but not for the phonemic discrimination test ($F_{1, 197} = .00$, $p = .99$, partial $\eta^2 = .00$, power = .05).

A closer look at the factor of hours of input can be helpful to understand what is happening between the two groups. Looking back to the scatterplots of total hours of study and test scores in Figures 1 and 2, it can be seen that the trend of the data for earlier learners was usually above the later starters in both tests when the total hours were somewhere between 1500 and 2000. In fact, a robust ANCOVA (20% trimmed means bootstrap- t ANCOVA) that compares points along the Loess line found statistical advantages for the earlier starters on the GJT at 1833 and 2000 hours, but a statistical advantage for later starters at 800 hours (other points did not show statistical differences). Given that none of the studies reported on in the García Mayo and García

Lecumberri (2003) volume examined learners who had had more than 600 hours of input, my results do not conflict with those of those researchers, but show that more hours of input may be necessary to begin to see differences between an early and later start. A robust ANCOVA examination of the phonemic discrimination scores found statistical advantages for the earlier starters at 1300, 1555, 1833 and 2000 hours (other points did not show statistical differences). Again, it seems to take a substantial accumulation of input to see advantages for an early start to begin to show.

3 Differences due to native speaking teachers

Given that there was a difference between earlier and later starter groups on the phonemic discrimination test, we might wonder whether having had a native-speaking English teacher (NEST) would have been helpful for the earlier starters in their phonemic ability. Of the 61 earlier starters, 34 reported that at least one of their teachers was a native speaker. Those with a NEST scored more highly on the phonemic discrimination test but had more variance (mean = 56.1; $sd = 16.1$) compared to non-NEST students (mean = 53.0; $sd = 12.8$). A t -test examining the difference between these two groups was not statistical, however ($t_{59,0} = .84, p = .40, \eta^2 = .01$; mean difference = 3.11 ± 7.41 , power = .23). A scatterplot in Figure 4 shows the distribution of phonemic discrimination test scores and the age that study began among the earlier starters. Looking at the smooth line, there seems to be a tendency for the very early starters (ages 4–6) to do much better on the test if they had NESTs, but the numbers are too small to test this intuition statistically. Further research with larger numbers of participants who started very young could help elucidate this area.

4 Attitudes towards language learning

My third research question was whether there would be any differences in attitudes towards language learning between the earlier and later starters. On the question of learning languages in general, the earlier starters had higher scores than the later starters on the 10-point scale (for scores, see Table 1). A t -test revealed that this difference was nearly statistical ($t_{197} = -1.9, p = .06, \eta^2 = .02$, power = .77) for the question of

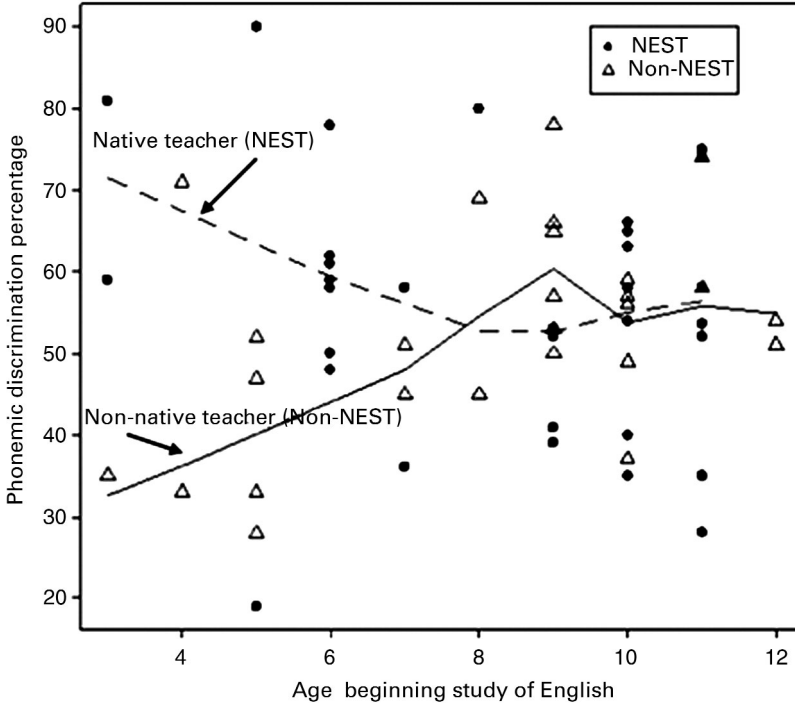


Figure 4 Scatterplot of beginning age for earlier starters and score on the phonemic discrimination test with division by whether or not teacher was a native English speaking teacher (NEST) or not (non-NEST)

learning languages in general as well as for learning English in particular ($t_{197} = -1.9, p = .06$). A correlation just among the earlier starters did not find any statistical relationship between age that study started and attitude towards learning English specifically ($r = -.16, n = 60, p = .22, \text{power} = .23$) or languages in general ($r = .04, n = 60, p = .74$).

VII Discussion

Contrary to many previous studies, this study found some modest effects for an early starting age in both grammatical and receptive phonological abilities. Age was not entirely separate from amount of input, however, and in fact interacted with it. Earlier starters gained an advantage on the grammaticality judgement test the younger they started their study of English, with differences in age accounting for 14% of the variation when controlling for total hours of input. An ANCOVA with the GJT

found total hours of input but not status as an earlier or later starter to be a statistical factor in explaining scores, and a robust ANCOVA further elucidated that earlier starters scored more highly than later starters on the GJT when they had a range of about 1600–2200 hours of input. In other words, morphosyntactic abilities could be enhanced by an early start, but only after a substantial amount of hours of input (this means from 6–8 hours a week if we calculate the 1600–2200 hours of input over 6 years and 44 weeks per year). When input levels became very high there were no statistical differences between the groups.

Earlier starters as a group scored statistically better than later starters on a test of phonemic discrimination, but no clear advantages for a younger starting age among the earlier starters was seen. The difference between groups accounted for about 3% of the variation among the earlier-starting and later-starting groups. There were no clear effects for having a native-speaking English teacher. Although the data graphics seemed to indicate that having a native speaker might be most helpful at quite young ages (4–6), the number of participants was too small to ascertain this statistically. A robust ANCOVA indicated an interaction between amount of input and group affiliation. It found that earlier starters held an advantage over later starters when they had a range of about 1200–2200 hours of input.

The fact that this study found some modest advantages to an earlier start in studying a classroom-instructed language, even with minimal input, can probably largely be attributed to the fact that learners in this study had larger amounts of total input than previous studies. Even though by definition the study was limited to classroom instruction of no more than 4 hours a week, participants reported significant amounts of homework and study outside the classroom that added to their total hours of study, and effects for being in the earlier starter group were not seen until the amount of input was fairly large. In fact, at least for the GJT, later starters scored statistically higher at 800 hours of input than earlier starters. This would fit in with the results from the studies in the García Mayo and García Lecumberri collection, since students in these studies were tested after at most 600 hours of classroom exposure (although it is not known whether these students would have had substantial amounts of homework and out-of-school classroom instruction in English as well).

We may ask why age differences would not become apparently advantageous until individuals had acquired a significant amount of input. One answer may be statistical. Looking at the scatterplots in Figures 1 and 2, it is clear that the majority of participants fall in the middle range (1000–2500) for total hours of input. Therefore, there may simply be more power to find differences in this range because there is a larger n . Given this line of reasoning, we might expect to find an advantage for earlier starters at every input level, if the group sizes were large enough. However, previous studies such as Burstall *et al.* (1974) have had quite large group sizes ($n = 5000+$) and still did not find many advantages for earlier starters.

Another answer may be that increasing amounts of input allow a more coherent language system to form, and as this occurs then advantages are able to more easily be demonstrated. This may be easiest to understand for the morphosyntax. For example, with very small amounts of exposure grammatical knowledge may still be fragmentary and tests of explicit judgement may favour those who learned a language in a more fragmentary, discrete-grammar-point type of way when they were older. Thus, the accumulation of many hours of input for the earlier starters may have finally allowed them to use their more integrated knowledge in an effective way. For phonemic discrimination, one might expect that the ability to discriminate phonemes of the language would not depend so much on an integrated knowledge of the entire language, but studies of Japanese learners of English have shown that productive pronunciation can definitely improve with more exposure to the language (Riney and Flege, 1998), so receptive phonological abilities may take some time to develop as well.

The effect sizes of the correlations and t -tests show that age plays only a modest explanative role, explaining from 3–14% of the variance in scores. Given the findings cited above that the ability to sound nativelike in phonology ends at an earlier age than the ability to do well on grammaticality judgement tests, we might have expected to find a stronger role for age in phonology. One possible reason for the discrepancy may be that there was a very wide range of exposure to nativelike phonology within this group, as opposed to what would be found for immigrants in an immersion situation. In other words, although we can be fairly sure that immigrants will be exposed to a variety of nativelike

accents and an abundance of spoken language, this is not clearly the case for children studying a foreign language. On the other hand, it is more likely that any input in English, whether spoken or written, would have contained abundant amounts of nativelike grammatical information, so the grammar is less likely to be affected by variations in quality and quantity of available spoken input than the phonology. Further research in the effects of early exposure to a language should try to better detail the type and amount of nativelike phonological input that learners received, and try to gather larger groups of learners who began at a very young age, where nativelike input may provide the most benefit.

As for attitudes towards learning languages in general or English specifically, which was the only advantage that earlier starters in García Mayo and García Lecumberri (2003) volume were found to have, there was a strong trend towards a statistical difference when looking at group differences, with earlier starters tending to have more positive attitudes towards studying languages in general and English in particular.

VIII Conclusions

This study was conducted to ascertain whether there would be any relationship between starting age and scores on a phonemic and morphosyntactic measure in a situation of minimal input. Contrary to predictions that age only plays a role in naturalistic or immersion environments, the present study found evidence that a younger starting age makes a modest difference to both phonological and basic morphosyntactic abilities, even in a situation of minimal input. The additional factor of input that was included in this study may be crucial in understanding the divergence of results from previous studies. These studies did not test participants with large amounts of input, and this study showed that advantages only emerged when total input increased. Contrary to what a lay person may think, age does not confer a 'magical' ability to learn a second language quickly and natively in a situation of minimal input. However, age does seem to play a non-negligible role in improving second language acquisition, given that language learners receive enough input. Starting to study a language at a younger age is

one way to ensure larger amounts of language input, so the present experiment overall finds a beneficial effect for starting to study a language at a younger age, even when input is only minimal.

Recommendations for educators and policy-makers

Because this study is in the minority of studies in that it found a positive effect for a younger starting age in a situation of minimal input, readers should be cautious in interpreting it too broadly. However, my recommendation would be to begin foreign language study as young as possible, with as many hours of input as are possible. I agree with Lee (1977) that few other school subjects, such as mathematics or reading, have been held up to the scrutiny that language learning has as to whether an earlier start is more effective than a later start. It is likely that older students are more effective at learning almost any kind of cognitive endeavour than younger students, but this does not stop us from putting young children in school and teaching them these subjects. With evidence of even modest benefits to a younger start for foreign language it then seems worthwhile to include this subject in the curriculum, especially as most studies concur that positive attitudes towards language learning develop with an early start. In fact, better morphosyntactic and phonemic discrimination ability may not be the only benefit that earlier learners receive. Other factors not tested for here – such as an ability to produce nativelike phonemes and morphosyntax, an increased vocabulary, more nativelike intonation, better speaking fluency or so on – may also result from an earlier start.

Acknowledgements

I would like to thank the students of Kyushu University who participated in this study, and the English department at that university for financial help provided to conduct this study. I benefited greatly from the comments of Greg Mizera as well as several anonymous *Second Language Research* reviewers, but of course any errors still found here are my own. Richard Herrington deserves special recognition for being an important influence on my statistics. An earlier version of this article was presented at the 2005 Second Language Research Forum conference.

IX References

- Bernstein, D., Roy, E., Srull, T. and Wickens, C.** 1988: *Psychology*. Boston, MA: Houghton Mifflin.
- Bialystok, E. and Miller, B.** 1999: The problem of age in second-language acquisition: influences from language, structure, and task. *Bilingualism: Language and Cognition* 2, 127–45.
- Birdsong, D.** 2005: Interpreting age effects in second language acquisition. In Kroll, J. and de Groot, A.M.B., editors, *Handbook of bilingualism: psycholinguistic approaches*. New York: Oxford University Press, 109–27.
- Bley-Vroman, R.** 1988: The fundamental character of foreign language learning. In Rutherford, W. and Sharwood-Smith, M., editors, *Grammar and second language teaching: a book of readings*. New York: Newbury House, 19–30.
- Brown, J.D.** 1995: English language entrance examinations in Japan: problems and solutions. Unpublished paper presented at the Proceedings of the JALT 1995 Conference.
- Buckby, M.** 1976: Is primary French really in the balance? *Modern Language Journal* 60, 340–46.
- Burstall, C.** 1977: Primary French in the balance. *Foreign Language Annals* 10, 245–54.
- Burstall, C., Jamieson, M., Cohen, S. and Hargreaves, M.** 1974: *Primary French in the balance*. Windsor: NFER Publishing.
- Butler, Y.G.** 2005: Comparative perspectives towards communicative activities among elementary school teachers in South Korea, Japan and Taiwan. *Language Teaching Research* 9, 423–46.
- Carroll, J.B. and Sapon, S.** 1959: *Modern language aptitude test: form A*. New York: The Psychological Corporation.
- Cenoz, J.** 2003: The influence of age on the acquisition of English: general proficiency, attitudes and code-mixing. In Garcia Mayo, M. and Garcia Lecumberri, M., editors, *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters, 77–93.
- Chaudron, C.** 2003: Data collection in SLA research. In Doughty, C. and Long, M.H., editors, *The handbook of second language acquisition*. Malden, MA: Blackwell, 762–828.
- Cohen, J.** 1988: *Statistical power analysis for the behavioral sciences*. Newbury Park, CA: Sage.
- DeKeyser, R.M.** 2000: The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22, 499–533.
- 2003: Implicit and explicit learning. In Doughty, C. and Long, M.H., editors, *The handbook of second language acquisition*. Malden, MA: Blackwell, 313–48.

- DeKeyser, R.M. and Larson-Hall, J.** 2005: What does the Critical Period really mean? In Kroll, J. and de Groot, A.M.B., editors, *Handbook of bilingualism: psycholinguistic approaches*. New York: Oxford University Press, 88–108.
- Dörnyei, Z. and Skehan, P.** 2003: Individual differences in second language learning. In Doughty, C. and Long, M.H., editors, *The handbook of second language acquisition*. Malden, MA: Blackwell, 589–630.
- Flege, J.E., Yeni-Komshian, G. and Liu, S.** 1999: Age constraints on second-language acquisition. *Journal of Memory and Language* 41, 78–104.
- García Lecumberri, M. and Gallardo, F.** 2003: English FL sounds in school learners of different ages. In García Mayo, M. and García Lecumberri, M., editors, *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters, 115–35.
- García Mayo, M.** 2003: Age, length of exposure and grammaticality judgments in the acquisition of English as a foreign language. In García Mayo, M. and García Lecumberri, M., editors, *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters, 94–114.
- García Mayo, M. and García Lecumberri, M.,** editors, 2003: *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.
- Gardner, R.C.** 1985: *Social psychology and second language learning: the role of attitudes and motivation*. London: Edward Arnold.
- Gass, S.** 1994: The reliability of second-language grammaticality judgments. In Tarone, E., Gass, S. and Cohen, A., editors, *Research methods in second language acquisition*. Hillsdale, NJ: Lawrence Erlbaum, 303–22.
- Hughes, H.J.** 1999: Cultivating the walled garden: English in Japan. *English Studies* 6, 556–68.
- Hyltenstam, K. and Abrahamsson, N.** 2003: Maturation constraints in SLA. In Doughty, C. and Long, M.H., editors, *The handbook of second language acquisition*. Malden, MA: Blackwell, 539–88.
- Jia, G.X., Aaronson, D. and Wu, Y.** 2002: Long-term language attainment of bilingual immigrants: predictive variables and language group differences. *Applied Psycholinguistics* 23, 599–621.
- Johnson, J.S. and Newport, E.L.** 1989: Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21, 60–99.
- Koike, I. and Tanaka, H.** 1995: English in foreign language education policy in Japan: toward the twenty-first century. *World Englishes* 14, 13–25.
- Lasagabaster, D. and Doiz, A.** 2003: Maturation constraints on foreign-language written production. In García Mayo, M. and García Lecumberri, M., editors, *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters, 136–60.
- Lederberg, A.R. and Spencer, P.E.** 2005: Critical periods in the acquisition of lexical skills: evidence from deaf individuals. In Fletcher, P. and

- Miller, J.F., editors, *Developmental theory and language disorders*. Philadelphia, PA: John Benjamins, 121–45.
- Lee, W.R.** 1977: For and against an early start. *Foreign Language Annals* 10, 263–70.
- Leow, R.P.** 1996: Grammaticality judgment tasks and second-language development. *Georgetown University Round Table on Languages and Linguistics*, 126–39.
- Lightbown, P.M.** 2000: Classroom second language acquisition research and second language teaching. *Applied Linguistics* 21, 431–82.
- Mandell, P.B.** 1999: On the reliability of grammatical judgement tests in second language acquisition research. *Second Language Research* 15, 73–99.
- Maronna, R.A., Martin, R.D. and Yohai, V.J.** 2006: *Robust statistics: theory and methods*. Hoboken, NJ: Wiley.
- Maxwell, S.E. and Delaney, H.D.** 2004: *Designing experiments and analyzing data: a model comparison perspective*. 2nd edition. Mahwah, NJ: LEA.
- Mayberry, R.I. and Eichen, E.B.** 1991: The long-lasting advantage of learning sign language in childhood: another look at the critical period for language acquisition. *Journal of Memory and Language* 30, 486–512.
- Morrow, P.R.** 1987: The users and uses of English in Japan. *World Englishes* 6, 49–62.
- Moyer, A.** 2004: *Age, accent and experience in second language acquisition*. Clevedon: Multilingual Matters.
- Muñoz, C.** 2001: Factores escolares e individuales en el aprendizaje formal de un idioma extranjero [Scholarly and individual factors in the formal comprehension of a foreign language]. *Estudios de Linguística*, supplement 1, 249–70.
- 2003: Variation in oral skills development and age of onset. In García Mayo, M. and García Lecumberri, M., editors, *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters, 161–81.
- Muñoz, C. and Tragant, E.** 2001: Motivation and attitudes towards L2: some effects of age and instruction. *Eurosla Yearbook* 1, 211–24.
- Oller, J. and Nagato, N.** 1974: The long-term effect of FLES: an experiment. *Modern Language Journal* 58, 15–19.
- Oyama, S.** 1976: A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research* 5, 261–83.
- Patkowski, M.S.** 1994: The critical period hypothesis and interlanguage phonology. In Yavas, M., editor, *First and second language phonology*. San Diego, CA: Singular, 205–21.
- Riney, T.J. and Flege, J.E.** 1998: Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition* 20, 213–43.
- Sasaki, M.** 1996: *Second language proficiency, foreign language aptitude, and intelligence: quantitative and qualitative analyses*. New York: Peter Lang.

- Sorace, A.** 1996: The use of acceptability judgments in second language acquisition research. In Ritchie and W.C., Bhatia, T.K., editors, *Handbook of second language acquisition*. San Diego, CA: Academic Press, 375–409.
- Tversky, A. and Kahneman, D.** 1971: Belief in the law of small numbers. *Psychological Bulletin* 76, 105–10.
- Verba, S. Verba, S., Kelman, S., Orren, G.R., Miyaki, I., Watanuki, J., Kabashima, I. and Ferree, G.D.** 1987: Elites and the idea of equality: a comparison of Japan, Sweden, and the United States. Cambridge, MA: Harvard University Press.
- Victori, M. and Tragant, E.** 2003: Learner strategies: a cross-sectional and longitudinal study of primary and high-school EFL teachers. In García Mayo, M. and García Lecumberri, M., editors, *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters, 182–209.
- Wilcox, R.** 2001: *Fundamentals of modern statistical methods: substantially improving power and accuracy*. New York: Springer.
- Yamada, R.A., Tohkura, Y. and Kobayashi, N.** 1996: Effect of word familiarity on non-native phoneme perception: identification of English /r/, /l/, and /w/ by native speakers of Japanese. In James, A. and Leather, J., editors, *Second language speech*. New York: Mouton de Gruyter, 103–18.

Appendix 1 Sample of words used in the phonological test

read	red	rook	right	royal
lead	lead	look	light	loyal
weed	wed	wook	wight	woyal