Journal of Memory and Language **44**, 235–249 (2001)
doi:10.1006/jmla.2000.2750, available online at http://www.academicpress.com on IDEAL®

On the Evidence for Maturational Constraints in Second-Language Acquisition

David Birdsong and Michelle Molis

University of Texas at Austin

Second-language (L2) acquisition is generally thought to be constrained by maturational factors that circumscribe a critical period for nativelike attainment. Consistent with the maturational view are age effects among learners who begin L2 acquisition prior to, but not after, closure of the putative critical period. Also favoring the maturational account is the scarcity of late L2 learners at asymptote who perform like natives, and weak effects of native language–target language pairings. With Korean and Chinese learners of English, the experimental study of Johnson and Newport (1989) yielded just these types of evidence. Some subsequent studies do not support the critical period account of L2 acquisition constraints, however. Accordingly, we undertook a replication of Johnson and Newport (1989), using the exact methods and materials of the original experiment, and a sample of Spanish natives ($n = 61$). In keeping with recent research, L2 attainment negatively correlates with age of learning even if learning commences after the presumed end of the critical period. We also find modest evidence of nativelike attainment among late learners. Our data further suggest that the outcome of L2 acquisition may depend on L1–L2 pairings and L2 use.

Key Words: second-language acquisition; maturation; age effects; critical period; bilingualism; linguistic development.

The level of ultimate attainment in second-language (L2) acquisition is constrained to a significant degree by the age at which learning begins. An "earlier is better" rule of thumb captures the negative correlation between age of learning onset and eventual asymptotic performance. A considerable body of experimental data is consistent with this generalization. Most of these studies offer a maturational account of age effects and suggest that a critical period limits both primary-language (L1) acquisition acquisition and L2 acquisition (see Long, 1990, for a comprehensive review).

Several types of behavioral evidence would constitute support for a maturational view of the limits on L2 attainment. First, linguistic performance should correlate negatively with the age at which L2 learning begins. Importantly, this effect should be observed in cases where L2 learning begins before the mid to late teens, that is, prior to the end of maturation. Among late learners, conversely, correlations of learning onset with performance would suggest a role for mechanisms other than maturation. Second, there should be few if any late learners who perform in the range of native controls; this null result would suggest biological constraints. Finally, if limits on attainment are maturational in nature, then they should apply to L2 acquisition generally. We would expect, therefore, that critical period-type effects and near-zero incidence of nativelike attainment should be observed no matter what the pairing of L1 and L2.

A few years ago, Jacqueline Johnson proposed the replication part of this study as a collaborative effort with the first author. At present, Jackie is inactive in research, and she has passed along her results from approximately 25 subjects in addition to providing subject compensation funds. We acknowledge with gratitude her generosity and cooperation. Portions of this article were presented at the 17th Second Language Research Forum, the 22nd Boston University Language Development Conference, McGill University, Université de Montréal, Harvard University, the University of Alabama at Birmingham, and the University of Texas at Austin. We are grateful for comments from audiences at these presentations. We also thank Jim Flege, Susan Guion, Kevin Gregg, Mike Long, Stefka Marinova-Todd, Elissa Newport, Carlota Smith, Catherine Snow, and Lydia White for discussion of our results. The article has also benefited from the careful reading of the reviewers of this journal.

Address correspondence and reprint requests to David Birdsong, Dept. of French & Italian B7600, University of Texas, Austin, TX 78712. Fax: (512) 471-8492. E-mail: birdsong@ccwf.cc.utexas.edu.

0749-596X/01 $35.00

The influential study of Johnson and Newport (1989), henceforth J&N89, proposed a maturational model of L2 attainment based on just these kinds of evidence. Results of their grammaticality judgment task, administered to a group of Chinese and Korean learners of English, were in line with the first two types of evidence. Further, the authors argued that their results should generalize to other L1–L2 contexts, and they alluded to work in progress that supported their contention. The findings and interpretations of J&N89 have been widely accepted, and the critical period account[1] of the limits of L2 acquisition has been promoted in mainstream L2 acquisition texts (e.g., Gass & Selinker, 1994; Towell & Hawkins, 1994), in a highly regarded review of the literature on maturational effects (Long, 1990), and in popular works on language acquisition such as Pinker (1994).

Since the publication of J&N89, some studies have further advanced the maturational view of limits of attainment in the L2. These include position papers (e.g., Eubank & Gregg, 1999), behavioral evidence (e.g., Johnson & Newport, 1991), evidence from brain imaging (e.g., Weber-Fox & Neville, 1999), and simulations of loss of language-learning capacity in evolutionary models (e.g., Hurford & Kirby, 1999) and in connectionist models (e.g., Marchman, 1993). For example, Shim (1994) investigated the acquisition of English by Koreans who varied in age of L2 immersion. On the basis of speed and accuracy in grammaticality judgments, Shim concluded that exposure to the L2 before the age of 5 assures nativelike performance on items representing Subjacency and Anaphoric Binding. Age effects were observed among learners whose English acquisition began between the ages of 6 and 11. For late learners, however, nativelike performance is not achieved, and "L2 acquisition occurs in a non-principled manner" (Shim, 1994, p. iv).

Other research has yielded results that do not support a maturational account. For example, Bialystok and Hakuta (1994, 1999), Birdsong (1992), Flege (1999), and others have found postmaturational age effects. Studies such as Bongaerts (1999), Cranshaw (1997), and van Wuijtswinkel (1994) have attested significant numbers of late learners who perform like natives on various linguistic tasks. And several studies (e.g., Bongaerts, 1999; Cranshaw, 1997; Flege, 1999) suggest that results are not generalizable, as the incidence of nativelike attainment appears to depend on particular pairings of L1 and L2.

Bialystok and Miller (1999) studied the acquisition of L2 English morphosyntax by native speakers of Spanish and Chinese using items similar to those of J&N89. For both groups, there was a main effect for modality of stimulus presentation (oral versus written) which is not predicted under a biologically based account of L2 acquisition constraints. Across the range of ages of arrival tested, age and performance were negatively correlated; however, in the regression analyses there was no point of inflection that would suggest the offset of a critical period. Further, divergences in the two learner groups' performances suggested native language effects.

Flege, Yeni-Komshian, and Liu (1999) investigated end-state proficiency in English pronunciation and morphosyntax by Korean natives of varying ages of arrival in the United States. With increasing Age of Arrival (AoA), there was a decline in pronunciation accuracy and performance on a 144-item subset of the J&N89 stimuli. When variables confounded with AoA were fac-

[1] Various uses of the term "critical period" appear to have in common the following assumptions: "First, learning during a critical period is assured, similar across individuals, normatively described, and probably governed primarily by endogenous factors. . . . Second, learning outside of the critical period is different in both form and success, especially in that it would be less certain and more erratic in its outcomes" (Bialystok & Hakuta, 1999, p. 164); see also Columbo (1982) and Bornstein (1989). Although many researchers associate the Critical Period Hypothesis for language acquisition with Lenneberg (1967), there are in fact multiple formulations. All share the features mentioned above but differ in terms of presumed neural or mental models of language acquisition. Customarily, these accounts are lumped under a single category name for the sake of convenience. Possible mechanisms underlying a critical period for L2 acquisition include maladaptive gain of processing capacity with cognitive maturation, loss of access to Universal Grammar, and dismantling of neural circuitry for language acquisition with increasing age (Birdsong, 1999, pp. 2–8). For the purposes of this study, we do not distinguish between a "critical period" and a weaker, more variable "sensitive period"; for discussion, see Eubank and Gregg (1999).

tored out, pronunciation appeared susceptible to age effects, although performance in morphosyntax did not. Participants' level of education in the United States predicted performance on rule-based areas of English morphosyntax, whereas their use of the target language correlated with performance on irregular features of English as well as with pronunciation accuracy.

It is in the context of such detailed findings that we reconsider the predictions of a maturational account of L2 attainment. A potential contribution to the scientific process would be to determine if the results of J&N89 are replicable. Partial replications of J&N89 have been conducted (Bialystok & Miller, 1999; Flege, Yeni-Komshian, & Liu, 1999; Jia, 1998; Johnson, 1992; van Wuitswinkel, 1994), and results have generally diverged from the original. However, because of procedural differences, or because of use of subsets or variations of the original stimuli, these findings cannot be compared directly to those of J&N89. More fruitful comparison may be achieved by a ceteris paribus replication of the J&N89 study. The present research employs the exact procedures and stimuli used by J&N89, with a sample of Spanish native speakers who learned L2 English.

THE JOHNSON AND NEWPORT (1989) STUDY

We begin by reviewing the design, methods, and findings of J&N89. The J&N89 participant sample consisted of 46 Korean and Chinese learners of English, representing *Age of Arrival* in the United States that ranged from 3 to 39 years. Half had moved to the United States by age 15 and were termed *Early Arrivals*; the *Late Arrivals* had left their home country at age 17 or thereafter. All had had at least 5 years' exposure to English and at least 3 years' uninterrupted residence in the United States prior to experimentation (average residence for Early Arrivals = 9.8 years; for Late Arrivals = 9.9 years). All subjects had had some schooling in the United States, and all were university faculty or students.

Participants were asked to give binary acceptability judgments of 276 randomly ordered English sentences. Each item was presented twice on audio tape, with a 1- to 2-s pause between the first and second readings and a similar interval between items. The sentences exemplified various surface features of English morphology and syntax. Twelve types of structures were represented, including present progressive, particle movement, and past-tense morphology. Approximately half the items were grammatical, the remainder being minimally varying ungrammatical counterparts. Sample items include the following:

> Every Friday our neighbor washes her car.
> *Every Friday our neighbor wash her car.
>
> Two mice ran into the house this morning.
> *Two mouses ran into the house this morning.
>
> The horse jumped over the fence yesterday.
> *The horse jumped the fence over yesterday.

In addition to carrying out the grammaticality judgment task, participants supplied information about their language background, attitudes about the United States, and motivation to learn English.

J&N89 reported the distribution of subject performances on the judgment task, plotted against their AoA. Prior to the presumed offset of a critical period for learning, the distribution of accuracy was represented by a linear function, with a decline of performance starting at AoA of 8–10 years and ending at an AoA of 15 years. In contrast, the distribution of scores was essentially random for subjects whose AoA surpassed the end of the putative critical period (Fig. 1).

Figure 1A shows a monotonic performance decrement as subjects' AoA increases from approximately 8 to 16 years. Among Early Arrivals, AoA was strongly predictive of results (r = –.87). However, among subjects whose AoA is between 17 and 39 (Fig. 1B), the age factor was not predictive, as indicated by a correlation coefficient approaching zero (r = –.16) and as suggested visually by the random dispersion of scores. In sum, after AoA of 7 years, and up to the end of maturation, AoA predicted L2 end-state performance. However, after maturation, AoA was no longer predictive. Thus, J&N89's results suggested that the L2 asymptote is determined not by a general age effect operating across the

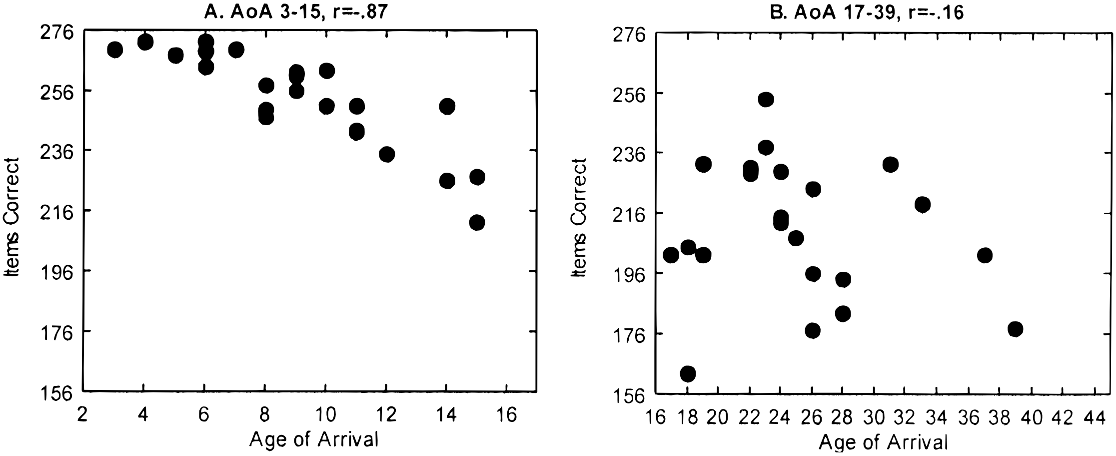

FIG. 1. The age function in J&N89, Early Arrivals (A), and Late Arrivals (B). In this and Figs. 2 and 3, points represent number of items correct for individual participants.

life span, but by neurocognitive developmental factors early on. J&N89 articulate this position:

> If the explanation for late learners' poorer performance relates to maturation, performance should not continue to decline over age, for presumably there are not many important maturational differences between, for example, the brain of a 17-year old and the brain of a 27-year old. Instead, there should be a consistent decline in performance over age for those exposed to the language before puberty, but no systematic relationship to age of exposure, and a leveling off of ultimate performance, among those exposed to the language after puberty. This is precisely what was found. (p. 79)

Also consistent with a maturational account were the disparate levels of task performance attained by early- versus late-arriving subjects in J&N89. The mean score of their subjects with the earliest AoA (3–7 years) was 269.3, with scores ranging from 272 to 264. These results are virtually identical to those of native controls, whose mean was 268.8, with a range of 275 to 265. Yet not one of the 23 Late Arrivals performed within the native range of scores (Late Arrivals range = 254–163; $M = 210.3$). This apparent strength of maturational constraints led J&N89 to conclude that nativelikeness is out of reach for late learners: "for adults, later age of acquisition determines that one will not become native or near-native in a language" (p. 81).

Finally, J&N89 presented a case for the generalizability of their results to learners whose native languages are not Korean or Chinese. (The proportion of Chinese to Korean natives is not given in J&N89, neither over all subjects nor by Age of Arrival groups. Thus the influence of the individual native languages on the outcome is not measurable.) They did acknowledge a limited role of the L1, recognizing that "certain aspects of the structure of one's first language are likely to have some effects on the learning of the second language" (p. 93). However, they also argued that the particular pattern of age-related results (Fig. 1) should extend beyond their sample: "We do not believe that the relationship found here between age of exposure and ultimate performance in the second language is unique to the circumstances where Chinese or Korean is the first language and English is the second" (p. 93). We assume that the relationship they refer to is the shape of the age function with accompanying lack of nativelikeness.

THE REPLICATION OF J&N89

With an emphasis on the types of evidence discussed above, we wished to determine if the results of J&N89 would replicate with a some-what larger sample and with native speakers of a language other than Korean or Chinese. In all other significant respects our study is identical to J&N89, thus permitting a high degree of comparability of results.

Participants

Our sample consisted of 61 native speakers of Spanish. Of these, 29 were Early Arrivals (AoA ≤ 16) and 32 were Late Arrivals (AoA ≥ 17). We attempted to ensure that learners were at asymptote by setting the mean length of residence in the United States at 10 years. Between early- and late-arriving groups, the length of residence is approximately equal. Thus, for Early Arrivals, the mean length of residence is 12.2 years compared to 10.5 years for Late Arrivals. According to a two-sample t test that assumes unequal variances, this difference is not significant [$t(55) = 1.12$, $p = .27$, two-tailed]. In addition, following J&N89, we determined that there was a nonsignificant correlation between AoA and years of residence, $r = -.12$, $p = .34$. Also, subject education characteristics are similar to those in J&N89. All had obtained at least a bachelor's degree, and all were students, faculty, or employees at major U.S. universities (Cornell University, University of Virginia, University of Illinois, or University of Texas at Austin).

Method and Materials

All materials and procedures were identical to those of J&N89. The original audiotape of the stimulus sentences was presented over the speaker of a Sharp RD 680AV in a quiet room. We supplemented the original biographical questionnaire with three items. One item asked participants to self-assess their fluency in English; the other two items elicited estimates of current use of Spanish and English. Another minor difference was that we considered responses to 274 tape-recorded items rather than 276. One pair of stimulus sentences from J&N89 was eliminated from the analyses because the putatively ungrammatical variant was judged by ourselves and other native English speakers to be acceptable in certain contexts. (The items in question are Right now Judy is shopping for a new party dress; *Right now Judy shops for a new party dress.)

Results and Discussion

For Spanish natives in the present study, Fig. 2 represents Age of Arrival plotted against accuracy on the acceptability judgment task.

For Early Arrivals (AoA ≤ 16), the correlation between judgment accuracy and AoA is not significant ($r = -.24$, $p = .22$). The distribution of scores in Fig. 2A suggests that this lack of

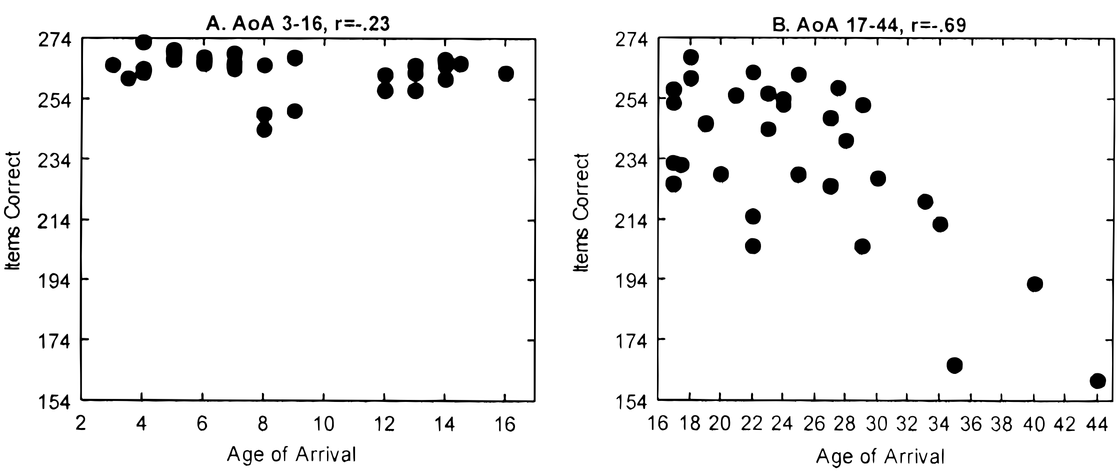

FIG. 2. The age function in the present study, Early Arrivals (A) and Late Arrivals (B).

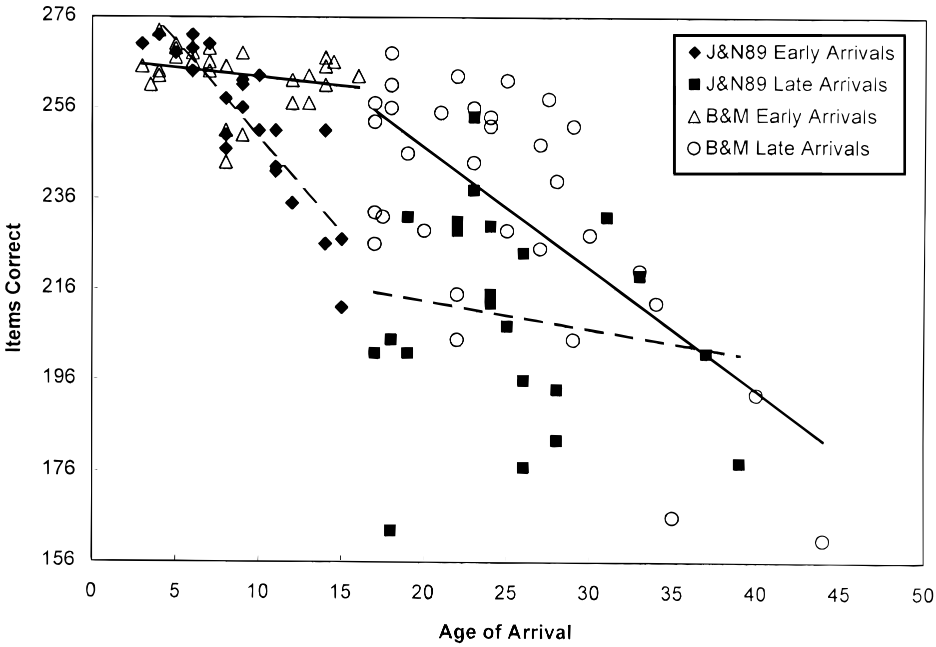

FIG. 3. Number of items correct as a function of Age of Arrival, current study, and J&N89. Regression lines (B&M: solid lines; J&N89: dashed lines) are provided for four data subgroups separated by study and by Early versus Late Arrivals. Cutoff age of AoA = 16.

correlation is the reflection of a ceiling effect, as scores for 28 of the 29 early-arriving subjects were ≥ 250. In contrast, for subjects whose AoA is ≥ 17, i.e., after the end of maturation, there is a strong age effect ($r = -.69$, $p < .0001$)[2]; see Fig. 2B. This result is to be contrasted with Fig. 1B, where AoA of the late-arriving J&N89 subjects did not correlate with performance.

A comparison of the current results with those of J&N89 reveals a significant difference in the regression slopes both before the cutoff age of 17 stipulated by J&N89 [$t(51) = 6.43$, $p < .01$] and after that age [$t(51) = 2.05$, $p < .05$]. Figure 3 displays the data and regression lines for the two studies.

[2] On the possibility that outliers in the distribution of Late Arrivals' scores skewed the results of this correlation, we reanalyzed the data, first removing the scores of the two Latest Arrivals, then dropping the scores of the three Latest Arrivals. The resulting values remained significant: $r = -.51$, $p = .004$ for the first reanalysis and $r = -.36$, $p = .05$ for the second. Similarly, from the J&N89 distribution of Late Arrival scores we removed the lowest score. Here, the result of the reanalysis approaches significance: $r = -.33$, $p < .14$.

Both J&N89 and the present study attest a strong age effect across the full range of AoA: J&N89 $r = -.77$, $p < .01$; present study $r = -.77$, $p < .0001$. (The identical coefficients do not imply similar slopes of the overall linear regressions; assuming fixed intercepts at the maximum possible score—276 for J&N89 and 274 for the present study—the slopes differ significantly [$t(105) = 5.26$, $p < .01$]. Age effects across the full AoA range are corroborated in other studies (e.g., Flege, 1999; Oyama, 1973) and are not consistent with a putative cessation of age effects at the end of maturation.

Table 1 displays distribution of performance over AoA in both studies. There is an apparent difference in onset of age effects between the two studies. For J&N89, an effect of AoA was found for subjects arriving between 8 and 10 years of age. In the present study, it is only at AoA ≥ 17 that a significant decline in performance is observed (t tests with the Bonferroni correction reveal no significant performance differences between the 3- to 7-, 8- to 10-, and 11- to 16-year AoA groups). Below we

TABLE 1

Grammaticality Judgment Performance

	Group				
	Natives ($n = 23$)	3–7 ($n = 7$)	8–10 ($n = 8$)	11–15 ($n = 8$)	17–39 ($n = 23$)
			J&N89[a]		
Mean	268.8	269.3	256.0	235.9	210.3
Standard deviation	2.9	2.8	6.0	13.6	22.8
Range	275–265	272–264	263–247	251–212	254–163
Errors	1–11	4–12	13–64	25–64	22–113
		3–7 ($n = 14$)	8–10 ($n = 6$)	11–16 ($n = 9$)	17–44 ($n = 32$)
			Present Study[b]		
Mean		266.7	256.8	262.3	234.5
Standard deviation		3.2	10.3	3.6	26.9
Range		273–261	268–244	267–257	268–161
Errors		1–13	6–30	7–17	6–113

[a] Nonnatives are grouped by Age of Arrival; maximum score = 276.
[b] Participants are grouped by Age of Arrival; maximum score = 274.

address statistical differences in onset timing between the two studies.

Bialystok and Hakuta (1994, p. 69) reanalyzed the age function in J&N89. Using a model that assumed no particular shape of the regression line, they, like J&N89, found that there were indeed two linear relationships. However, the articulation "elbow" at the disjuncture of these lines appeared to occur at AoA 20, considerably later than the AoA 16, which was the cutoff point for J&N89's analysis of Early versus Late Arrivals. When Bialystok and Hakuta extended the cutoff age to 20, the correlation coefficient for early arrivals remained at $r = -.87$, but the magnitude of the linear coefficient for late arrivals rose to statistical significance ($r = -.50$, $p < .05$). This was seen as undermining J&N89's claim of no age effects past maturation. With respect to the present data, when the cutoff age is extended to AoA = 20, the correlation is significant for Early Arrivals ($r = -.56$, $p < .001$) as well as for Late Arrivals ($r = -.71$, $p < .001$). Again assuming a cutoff at AoA = 20 for our data and those of J&N89, we find significant differences across the data sets between the regression slopes for Early Arrivals [$t(66) = 6.70$, $p < .01$] as well as for Late Arrivals [$t(40) = 2.67$, $p < .05$].

We performed an additional analysis to further test J&N89's assumption that their data are most appropriately characterized by two linear functions separated by a cutoff age corresponding to the end of a putative critical period. First we tested the fit of a piecewise regression relative to a single linear function, with the intercepts fixed at the highest possible scores. The inflection point between the two linear segments was included as a free parameter. The piecewise regression provided a significant improvement in the fit to the J&N89 data [$F(2, 42) = 3.25$, $p = .048$] and placed the cutoff age at 18 years. However, subsequent tests reveal that a cutoff of AoA = 18 does not produce a better fit to the data than the cutoffs proposed by J&N89 [AoA = 15; $F(1, 42) = 0.67$, $p = .42$] or by Bialystok and Hakuta (1994) [AoA = 20; $F(1, 42) = 1.29$, $p = .26$].

The same set of analyses was carried out on the current set of data. The fit to the data was significantly improved for the piecewise regression with the inflection point included as a free parameter, relative to a single linear function [$F(2, 59) = 7.17$, $p < .01$]. It is interesting to note, however, that the cutoff age provided by this analysis was 27.5 years. We made additional comparisons to the specific cutoff ages proposed by J&N89 and by Bialystok and Hakuta (1994). The fit to the data provided by the cutoff age of 27.5 represents an improvement over the J&N89 cutoff of AoA = 15 [$F(1, 59) = 6.92$, $p < .05$] but not over the AoA = 20 [$F(1, 59) = 3.69$, $p = .06$] suggested by Bialystok and Hakuta (1994).

To summarize, we compared our results with those of J&N89, using a variety of statistical techniques. The studies are significantly different in terms of the onset of age effects; the correlation of scores with early and late AoA; and the regression slopes for Early Arrivals, Late Arrivals, and over all participants.

Note that with later AoA both studies found increasing variability in performance. This progression toward more pronounced intersubject differences is suggested visually in Fig. 3 and by the standard deviation figures in Table 1. As J&N89 argue (pp. 96–97), greater individual differences after maturation are consistent with the Critical Period Hypothesis. However, the dispersion of Late Arrivals' scores in our study is relatively minor. In other words, the age effect persists across all AoAs despite the tendency toward increasing individual differences. The strength of the correlation of AoA and performance for our Late Arrivals is remarkable given the observed widening of interindividual differences—across many domains of cognition and performance—as age increases (e.g., Flavell, 1970; Schaie, 1994; Weinert & Perner, 1996).

Thus, in our study age of acquisition predicts performance both postmaturationally and over the entire span of AoA. Consider the possibility that these results are artifactual. For example, the ceiling effect, which extends across our three subgroups of Early Arrivals, might result from items that are not sufficiently challenging. Recall, however, that our replication used the original stimuli; these items were indisputably challenging for Chinese and Korean subjects with AoA ≥ 7. We note too that, for Late Arrivals in our study as well as in J&N89, scores are plainly depressed and decline with increased AoA.

Another candidate source of divergent onsets of age effects is a potential confound of AoA with Age of First Exposure to English prior to coming to the United States. A prolonged and rich exposure to the target language, prior to immersion, could conceivably retard the effect of age on ultimate attainment. (For the sake of argument, we assume the best case scenario of rich and prolonged exposure, although our data do not quantify these dimensions.) On this variable it is impossible to make a direct comparison of the two studies, since we do not have the raw Age of First Exposure data from J&N89. For their subjects who had had classroom instruction, J&N89 do perform the correlation of age of exposure and performance; the correlation is not significant, $p > .05$. Among the relevant subjects in our study (i.e., those with AoA ≤ 16), the correlation between age of first exposure and performance just misses significance, $p = .053$; see discussion of biographical variables below. This marginal result suggests that further investigation of a possible confound is warranted, perhaps involving comparisons of groups matched for Age of Exposure and varying in Age of Arrival (Flege et al., 1999).

In a finer grained analysis, we consider the possibility that *subgroups* of Early Arrival had had abnormally early first exposure. In Table 2 we break down our Early Arrivals into the AoA subgroups used by J&N89 and display relevant age of first exposure data. In the range data note that, for the 3- to 7- and 8- to 10-year-old AoA subgroups, the age of first exposure is essentially equatable with the age at which they arrived in the United States. It is only in the 11- to 16-year-old AoA subgroup that first exposure

TABLE 2

Ages of First Exposure to English, Early Arrivals, Present Study

	Age of arrival group (years)		
	3–7 ($n = 14$)	8–10 ($n = 6$)	11–16 ($n = 8$)
Mean age of first exposure	5.1	8.3	9.2
Standard deviation	1.2	0.5	3.7
Range	2.5–7	8–9	4–14

Note. $n = 28$ (one participant did not respond).

precedes immigration. Early exposure may indeed have given these learners an advantage, with the consequence that AoA is not the (only) age factor that is implicated in critical period-type effects. However, for the individual early-arriving subgroups, including the 11- to 16-year-old AoA subgroup, correlations of age of first exposure with performance do not reach significance. (Below, we discuss the separate issue of age of first exposure among Late Arrivals.)

We also examined the biographical profiles of our participants to see if there were experiential factors that might be related to the timing of age effects. With AoA = 8–10 and AoA = 11–16 subgroups considered separately, we correlated each surveyed biographical factor (see footnote to Table 3) with performance on the judgment task. Results were not revealing, however, as none of the correlations reached significance.

Finally, to arrive at the L2 acquisition asymptote—whether this end state be mastery or something short of mastery—it may be that fewer years of exposure to English are required of Spanish speakers than of Korean or Chinese speakers (Odlin, 1989, pp. 38–42). If this is the case, one could imagine any number of resultant differences in eventual attainment, including the timing of age effects. However, this scenario begs the question of why it should take Spanish speakers relatively less time to reach the end state. It also begs the larger question of whether onset-timing differences as a function of the L1 and L2 can be accommodated under a maturational model of L2 acquisition constraints.

TABLE 3

Correlations of Accuracy with Biographical Variables, Present Study

	Biographical variable[a]					
	Motivation 1st	% English exp 1st	% Spanish exp 1st	Self-consciousness	% English use now	% Spanish use now
Early Arrivals (AoA ≤ 16)	.235	.040	.040	.394*	.167	.186
Late Arrivals (AoA ≥ 17)	.052	.356*	–.356*	–.222	.401**	–.401**
All	.115	.039	–.039	–.226*	.446****	–.454****
	Importance	Fluency	Motivation	ID w/U.S.	English study	Age of 1st exposure
Early Arrivals (AoA ≤ 16)	–.131	.233	–.409*	.327*	.156	–.362*
Late Arrivals (AoA ≥ 17)	–.085	.130	.060	.050	.380**	–.451***
All	–.069	.630****	–.140	.169	.046	–.552****

[a] Definitions of biographical variables: motivation 1st: “In the first years of learning English, how motivated to learn were you?” (on scale of 1–5); % English exp. 1st: “When you first arrived in the U.S., how much English were you exposed to?” (percentage of time); % Spanish exp. 1st: “When you first arrived in the U.S., how much Spanish were you exposed to?” (percentage of time); self-consciousness: “Did you feel self-conscious while learning English in the U.S.?” (on scale of 1–5); % English use now: “How much of your present speaking and listening behavior is done in English?” (percentage of time); % Spanish use now: “How much of your present speaking and listening behavior is done in Spanish?” (percentage of time); importance: “How would you rate the importance of being able to speak English well?” (on scale of 1–5); fluency: “How fluent are you in English?” (on scale of 1–5); motivation: “How motivated are you (presently) to improve your English?” (on scale of 1–5); ID w/U.S.: “How strongly do you identify with the American culture?” (on scale of 1–5); English study: “How long did you take English classes?” (years); and age of 1st exposure: “How old were you when you first started learning English?” (years).

*$p < .1$.

**$p < .05$.

***$p < .01$.

****$p < .001$.

Clearly, the plausibility of the maturational account hinges in part on the age at which declines in attainment begin. If there are no age-related effects prior to the closure of the presumed critical period, then maturation cannot be isolated as the source of performance decrements. J&N89 (p. 96) acknowledge that the timing of age effects is a challenging issue for those working in the critical period paradigm, especially if age effects begin after maturation is complete. (For discussion of various onset and offset ages proposed in the literature, see Moyer, 1999, p. 100.)

Turning now to the question of nativelike attainment (see range and error data in Table 1), recall that J&N89 found no late-arriving subjects who scored in the range of natives or in the range of the earliest arrivals. This absence of nativelike attainment among late L2 learners is consistent with many studies (e.g., Coppieters, 1987; Oyama, 1976; Patkowski, 1980; and references in Long, 1990).

In contrast, other research with late L2 learners finds nativelike levels of attainment. For example, van Wuijtswinkel studied Dutch native speakers who judged the grammaticality of a subset of the Johnson and Newport (1989) items, along with an assortment of other syntactic structures in English. (These participants were residents of The Netherlands, not immigrants to an Anglophone country; age of exposure to English ≥ 12 years.) Eight of 26 participants in one group, and 7 of 8 participants in a more advanced group of learners, performed like natives. In a study of Francophone learners of English, White and Genesee (1996) found that many of their subjects who had had their first significant exposure to English after age 12 performed like natives on production and judgment tasks relating to items exemplifying constraints on wh-movement. On the grammaticality judgment task in Birdsong (1992), most of the 20 nonnatives performed within the native range. Birdsong (1997) explored Anglophones' acquisition of the idiosyncratic distribution of the pronominal clitic *se* in French intransitive predicates. Four of the 20 late learners performed at ≥ 94.8% accuracy on a grammaticality judgment task (mean accuracy for natives = 95.2%).

For phonetics and phonology—components of language thought to be most vulnerable to critical period effects (Scovel, 1988)—nativelike performance has been demonstrated in several studies by Bongaerts and his colleagues. In Bongaerts (1999), for example, samples of native English subjects and Dutch native speakers exposed to English after age 12 were rated blindly for authenticity. Five of 41 Dutch subjects received ratings within 1.2 standard deviations of the native English mean. In a similar study with French as the target language, 3 of 36 Dutch participants met this criterion. In a study of constraints on liaison in French, Birdsong (1997) found that 4 of 20 participants made no errors on a reading-aloud task.

Nativelike performance has also been observed by Cranshaw (1997); Juffs and Harrington (1995); Ioup, Boustagui, El Tigi, and Moselle (1994); Mayberry (1993); and White and Juffs (1997). In most studies where nativelike attainment is found, subjects who perform at nativelike levels comprise about 5–20% of the sample. As this is a significant proportion of an essentially random sample of late bilinguals (see discussion of selection criteria), nativelikeness is not an aberrant occurrence that could be dismissed as irrelevant to the study of L2 acquisition (cf. Bley-Vroman, 1989; Selinker, 1972). (It is likely that the folk wisdom of insignificant incidence of nativelikeness assumes the population of all those who have attempted to learn a foreign tongue rather than the relevant population who, like L1 learners, are essentially immersed in the ambient language.)

In the present study, one of our Late Arrivals correctly answered 268 items, a score that is within the range of natives sampled by J&N89. In addition, of the 32 participants arriving in the United States at age 17 or later, 3 had scores that were above 95% accuracy (260/274). Thirteen of our late-arriving participants had scores that were at or above 92% accuracy compared to only 1 such Late Arrival in J&N89. To what degree can this elevated performance be taken as evidence against a maturational account of the L2 acquisition end state? On the one hand, they meet Long's criterion for rejecting the Critical Period Hypothesis: "a single learner who began learning after the [critical period] closed and yet whose underlying linguistic knowledge . . . was shown to be indistinguishable from that of a monolingual native speaker would serve to refute the [Critical Period Hypothesis]" (p. 255). Yet they represent only modest evidence that late learners can reach nativelike levels of performance, and they fall short of the incidence of nativelikeness (up to 20% of the sample) found in other recent studies. In sum, our results lend themselves to either a half-full or half-empty interpretation; that is, as evidence of (near) nativelikeness or as evidence of depressed levels of performance relative to that of natives. Such findings do not contribute decisively to the issue of nativelikeness at the end state of L2 acquisition. Nevertheless, they are in line with the trend toward high levels of performance found in other research and are markedly distinct from those in J&N89.

We wish to point out that many factors could invalidate demonstrations of nativelike attainment by artificially elevating subjects' performance. One possible culprit is tasks and stimuli that do not probe the full grammar of the target language. For example, tests of morphosyntax that are limited to basic word-order patterns and common inflectional suffixes might fail to diagnose subjects with nonobvious linguistic deficits. (As suggested above, the J&N89 stimuli are not immune to this criticism. Recall, however, that for the Chinese and Korean sample studied by J&N89, the stimuli prove to be challenging. Those who would question the evidence for nativelike attainment in late L2 acquisition cannot, in the absence of principled argumentation, simultaneously accept the results for the Korean and Chinese sample and dismiss them for the Spanish sample.) However, several studies that find significant numbers of nativelike performers among late arrivals had deliberately constructed tests requiring knowledge of subtle, complex, and even obscure structures in the target language (e.g., Birdsong, 1992, 1997; Bongaerts, 1999; White & Genesee, 1996; van Wuijtswinkel, 1994).

Another possible source of elevated performance is careful screening of subjects. Some researchers (e.g., White & Genesee, 1996) went out of their way to identify the most advanced learners, as their goal was to determine the uppermost limits of acquisition. As suggested in the description of subjects, this is not the case in the present study. Moreover, in all other end-state studies conducted by the first author, subjects are not restricted to those who are self-assessed or independently determined to be "near natives"; rather, they merely satisfy a length-of-residence requirement and are college educated. Thus in Birdsong (1992) the residence requirement is ≥ 3 years of continuous residence, and in subsequent end-state studies he conducted the requirement is ≥ 5 years' residence. Within this population, the sampling is not biased toward the upper end of target-language mastery.

Finally, it is possible that nativelike late arrivals in the present study, prior to their definitive immigration to the United States had had early exposure to English. That is, late-arriving subjects with early exposure might have begun their learning prior to the closure of a presumed critical period. Looking closely at the biographies of our the 13 Late Arrivals who scored ≥ 92% on the task, we found little evidence that their ages of first exposure to English are associated with nativelike performance on the task. The mean first exposure of these Late Arrivals is 8.5 years as compared to the 6.89 years among all Early Arrivals. Among our 13 highest scoring Late Arrivals, some had indeed had early first exposure (range = 2.5–20 years), but those with the earliest exposure were not necessarily the best performers, as the correlation of age of first exposure with performance is not significant ($r = -.068$, $p = .83$). Further, the partial correlation of Late Arrivals' performance with (AoA minus first exposure) is significant, $r = -.648$ ($p < .01$). However, the difference in the mean performance of Late Arrivals with early first exposure (first exposure < 12 years old, $n = 20$) and those without (first exposure > 12 years old, $n = 12$) approaches significance [$F(1, 30) = 2.88$, $p = .10$].

Cook (1997) and Grosjean (1989) question whether nativelike behavior is appropriate data in bilingualism research. By definition, an L2 learner cannot be or become a native speaker, and thus it is pointless to hold out the monolingual native as a yardstick of success. Still, examples of nativelikeness challenge received views of late L2 acquisition, as they demonstrate that the upper limits of late L2 acquisition are not necessarily inferior to those of L1 acquisition (Mack, 1997).

We now revisit the question of generalizability of results. Under the argument advanced by J&N89, if a maturationally based critical period largely determines the level of ultimate attainment and the nature of the age function, then their findings should not be unique to the sampled population. Thus, the effect of L1–L2 pairing (a factor unrelated to maturation) should be minimal. This hypothesis was not supported in our replication study. With native speakers of Spanish we found that effects of age continue past the presumed end of maturation. Among our Spanish sample the onset of age effects differs from what J&N89 found for Chinese and Korean natives. Further, both prior to and after maturation, the slope of the age functions differs across the two data sets, suggesting variability in age effects as a function of participants' L1.

Our choice of L1 Spanish was based on several considerations. In L2 development short of the end state (e.g., initial settings of parameters, transitional stages in parameter resetting, subjects' intuitions for L2 linguistic markedness, and rate of attainment of L2 features), the influence of a given native tongue on specific areas of a given target language has been amply established (e.g., Eubank, 1994; Giacobbe, 1992; Kellerman & Sharwood Smith, 1986; Montrul, 2000; Odlin, 1989; Vainikka & Young-Scholten, 1996; cf. Epstein, Flynn, & Martohardjono, 1996, for a view of L2 acquisition that minimizes L1 influence). Spanish and English are related in many fundamental respects: their unmarked word order is SVO, the two vocabularies share many cognates, the grammars use both inflectional and adverbial means of marking tense/aspect distinctions, and so on. Under a parameter-setting view, there are relatively few parameters of Spanish that require resetting to English values. Arguably, these and other similarities might give native speakers of Spanish, relative to Korean or Chinese natives, at least some short-term advantages in rate of learning English (Odlin, 1989, and discussion above).

In the present study we are particularly concerned with the end state, not initial or intermediate stages of learning. Sorace (1993) argues that end-state L2 grammars differ qualitatively as a function of the L1. Indeed, empirical studies of different pairings of native and target languages show felicitous pairings of typologically similar L1s and L2s (e.g., Bialystok & Miller, 1999; Bongaerts, 1999; Cranshaw, 1997; Kellerman, 1995; Marinova-Todd, 1994; Sorace, 1993; van Wuijtswinkel, 1994). For example, with respect to the frequency of nativelikeness in L2 English, Cranshaw (1997) found among French natives a higher incidence of nativelike syntax and tense/aspect distinctions than among Chinese natives. Similarly, for English pronunciation, Bongaerts (1999) reports that Dutch natives at asymptote in English L2 are more likely to attain nativelikeness than French natives.

To what degree are divergent rates of nativelikeness and onset of age effects attributable to L1 influence and not to other factors? One approach to this question would be to carry out an additional replication of J&N89, using a larger sample of Korean and Chinese participants. Similarly, one might replicate J&N89 with an adequate sample of just Korean natives or just Chinese natives, as the two languages are sufficiently different from each other to warrant independent investigations.

L1 influence can also be identified at the level of structure; that is, by isolating L2 grammatical features which, under a contrastive analysis of the L1 and L2 grammars within a principled framework such as parameter setting, are predicted to be differentially mastered as a function of the L1–L2 pairing. This is the approach taken in the studies cited above. Bialystok and Miller (1999) and Kellerman (1995) note that to make L1-to-L2 causal inferences, the structural features of the target language must be exemplified unambiguously. However, in J&N89, the "rule types" do not constitute homogeneous categories. For example, within the set of items exemplifying wh-questions, three distinct syntactic phenomena are tested: the requirement of do-support (*What they sell at the corner store?), subject–auxiliary inversion (*When Sam will fix the car?), and choice of wh-word to reflect the semantics of the matrix verb (*Why did she put the book?). Consequently, for purposes of determining L1 influence, a rule-type-by-rule-type comparison of the two studies is not fruitful.

Finally, we consider the role of biographical variables in L2 attainment. Our data consist of answers to the J&N89 questionnaire, supplemented by items probing English vs Spanish use. Correlations are displayed in Table 3, with paraphrases of the questions given beneath the table.

Our results are broadly in line with those of J&N89. Assorted correlations reach significance, but none approaches the strength of the correlation between AoA and accuracy ($r = -.69$ for Late Arrivals and $r = -.77$ over all subjects).

Importantly, among our Late Arrivals and over all our subjects, amount of current English use is a strong predictor of performance. (One would expect identical correlation coefficients relating to percentage of English versus percentage of Spanish—at initial exposure and for current use—although one coefficient should be positive and the other negative. Indeed, this true for the Late Arrivals. However, for Early Arrivals and over all subjects, the figures for percentage of English use now and percentage of Spanish use now do not follow this pattern. This apparent discrepancy is explained by the fact that one of the Early Arrivals sometimes spoke a third language in addition to Spanish and English.) J&N89 do not have data on their participants' use of L2, but posit a weak influence of input relative to the effects of maturation. Thus, consistent with Flege, Frieda, and Nozawa (1997); Flege (1999); and Flege, Yeni-Komshian, and Liu (1999), our results suggest a nontrivial role for practice in determining the L2 acquisition end state.

We return once more to the possibility that biographical factors and AoA are confounded, thus reducing the predictive value of the latter variable. By partial correlation procedures both J&N89 and the present study find that AoA is the strongest and most reliable predictor of performance. For Late Arrivals in our study, partial correlations of AoA with performance, removing the various biographical factors, yield coefficients ranging from .619 to .800 (absolute values); by contrast, partials of biographical factors minus AoA give coefficients ranging in absolute value from .008 to .548. Recall in particular that for Late Arrivals the correlation of AoA with performance, partialing out first exposure, yields $r = -.648$ ($p < .01$). Conversely, we obtain $r = -.349$ (ns) for the correlation of (Late Arrivals' first exposure minus AoA) with performance. In correlations involving all subjects, comparable results were observed. For Early Arrivals, none of the partial correlations reached significance, as the performance of this group was at ceiling.[3]

CONCLUSIONS

We undertook the present study as a furtherance of J&N89 (see footnote on title page), whose scientific contribution to L2 acquisition theory cannot be overstated. Our investigation concentrated on three types of evidence that would favor a critical period account of L2 acquisition limits: age effects prior to but not after the end of maturation, near-zero incidence of nativelike levels of attainment, and the generalizability of these results across L1–L2 contexts. Our strict replication of J&N89 yielded evidence of postmaturational age effects and native language effects. Modest evidence of nativelike performance was also observed.

Researchers have argued these results would be evidence for falsification of the Critical Pe-

[3] We do not infer causality from this correlational evidence: relatively high English use could promote increased proficiency or be the result of it. For further discussion of maturation versus exogenous biographical factors, see Marinova-Todd, Marshall, and Snow (2000) and Moyer (1999). It is beyond the scope of this article to discuss the range of endogenous factors that might predict end-state attainment in L2 acquisition. Several studies have looked at relationships of aptitude, motivation, anxiety, cognitive style, and so on with rate of attainment or level of attainment short of the end state (e.g., Skehan, 1989; van Els, Bongaerts, Extra, van Os, & Janssen-van Dieten, 1984, pp. 103–125), but few have focused on the L2 acquisition asymptote (Klein, 1995). Interestingly, Schneiderman (1991) has found that various measures of language-learning aptitude (e.g., California Verbal Learning Test, Wechsler Memory Quotient, Modern Language Aptitude Test, Verbal IQ, and the Digit Symbol subtest of the Wechsler Adult Intelligence Test—Revised) do not correlate with L2 acquisition outcome.

riod Hypothesis as it relates to L2 acquisition. The postmaturational age-related effects we found are fairly robust and are grounds for refutation according to Pulvermüller and Schumann (1994), who maintain that "if the decrease in grammatical proficiency with greater age in postpuberty starters could be confirmed, the present [neurobiologically-based] proposal would have to be modified" (p. 723). On the matter of nativelikeness, our evidence exceeds Long's Popperian criterion for rejecting the Critical Period Hypothesis (Long, 1990, p. 255). Finally, extrapolating from J&N89's assertion of generalizability (p. 93), these criteria for rejection should apply across L1–L2 pairings.

Our approach was narrowed by adherence to the procedures, materials, and analyses of J&N89. Nevertheless, our results are consistent with those of less restrictive studies reviewed earlier. Further, our study and others find correlations of attainment with exogenous factors, such as amount of target language use. In the future, our understanding of the limits of L2 acquisition may be enhanced by additional replications of J&N89. We also look forward to finer grained analyses that disentangle confounded biographical variables and that relate experiential factors to principled linguistic distinctions such as regular versus irregular inflectional morphology (e.g., Flege et al., 1999).

REFERENCES

Bialystok, E., & Hakuta, K. (1994). *In other words: The science and psychology of second-language acquisition*. New York: Basic Books.

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 161–181). Mahwah, NJ: Erlbaum.

Bialystok, E., & Miller, B. (1999). The problem of age in second language acquisition: Influences from language, task, and structure. *Bilingualism: Language and Cognition*, **2,** 127–145.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, **68,** 706–755.

Birdsong, D. (1997, November). *Intransitivity and SE in French: Aspects of late L2 learnability*. Paper presented at the 22nd Boston University Child Language Development Conference.

Birdsong, D. (1999). Introduction: Whys and why nots of the Critical Period Hypothesis. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 1–22). Mahwah, NJ: Erlbaum.

Bongaerts, T. (1999). Ultimate attainment in foreign language pronunciation: The case of very advanced late foreign language learners. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 133–159). Mahwah, NJ: Erlbaum.

Bornstein, M. H. (1989). Sensitive periods in development: Structural characteristics and causal interpretations. *Psychological Bulletin*, **105,** 179–197.

Columbo, J. (1982). The critical period concept: Research, methodology, and theoretical issues. *Psychological Bulletin*, **91,** 260–275

Cook, V. J. (1997). Monolingual bias in second language acquisition research. *Revista Canaria de Estudios Ingleses*, **34,** 35–49.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, **63,** 544–573.

Cranshaw, A. (1997). *A study of Anglophone native and near-native linguistic and metalinguistic performance*. Unpublished Ph.D. dissertation, Université de Montréal.

Curtiss, S. (1988). Abnormal language acquisition and the modularity of language. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey* (Vol II, pp. 96–116). Cambridge, UK: Cambridge Univ. Press.

Epstein, S., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, **19,** 677–758.

Eubank, L. (1994). Optionality and the initial state. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 369–388). Amsterdam: Benjamins.

Eubank, L., & Gregg, K. R. (1999). Critical periods and (second) language acquisition: *Divide et impera*. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 65–99). Mahwah, NJ: Erlbaum.

Flavell, J. H. (1970). Cognitive changes in adulthood. In L. R. Goulet & P. Bates (Eds.), *Life-span developmental psychology* (pp. 247–253). New York: Academic Press.

Flege, J. E. (1999). Age of learning and second-language speech. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 101–131). Mahwah, NJ: Erlbaum.

Flege, J. E., Frieda, A. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, **25,** 169–186.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, **41,** 78–104.

Gass, S. M., & Selinker, L. (1994). *Second language acquisition: An introductory course*. Hillsdale, NJ: Erlbaum.

Giacobbe, J. (1992). A cognitive view of the role of the L1 in the L2 acquisition process. *Second Language Research*, **8,** 232–250.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, **1,** 131–149.

Hurford, J. R. & Kirby, S. (1999). Co-evolution of language size and the critical period. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 39–63). Mahwah, NJ: Erlbaum.

Jia, G. X. (1998). *Beyond brain maturation: The Critical Period Hypothesis in second language acquisition revisited*. Unpublished Ph.D. dissertation, New York University.

Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning*, **42,** 217–248.

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

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, **39,** 215–258.

Juffs, A. (1996). Semantics–syntax correspondences in second language acquisition. *Second Language Research*, **12,** 177–221.

Kellerman, E. (1995). Age before beauty: Johnson and Newport revisited. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of Interlanguage: Studies in honor of William Rutherford* (pp. 219–231). Amsterdam: Benjamins.

Kellerman, E., & Sharwood Smith, M. (1986). *Crosslinguistic influence in second language acquisition: An introduction*. New York: Pergamon.

Klein, W. (1995). Language acquisition at different ages. In D. Magnusson (Ed.), *The lifespan development of individuals: Behavioral, neurobiological, and psychosocial perspectives: A synthesis* (pp. 244–264). Cambridge, UK: Cambridge Univ. Press.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. London: Longman.

Lenneberg, E. H. (1976). *Biological foundations of language*. New York: Wiley.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, **12,** 251–285.

Mack, M. (1997). The monolingual native speaker: Not a norm, but still a necessity. *Studies in the Linguistic Sciences*, **27,** 113–146.

Marinova-Todd, S. H. (1994). *The critical period in second language acquisition: The case of gender*. Unpublished thesis, York University.

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three misconceptions about age and L2 learning. *TESOL Quarterly*, **34,** 9–34.

Montrul, S. (2000). Transitivity alternations in L2 acquisition: Toward a modular view of transfer. *Studies in Second Language Acquisition*, **22,** 229–273.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, **21,** 81–108.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge, UK: Cambridge Univ. Press.

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, **30,** 449–472.

Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: Morrow.

Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language Learning*, **44,** 681–734.

Schaie, K. W. (1994). Developmental designs revisited. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological considerations* (pp. 45–64). Hillsdale, NJ: Erlbaum.

Schneiderman, E. (1991). *Some neuropsychological characteristics of talented adult language learners*. Paper presented at the Second Language Research Forum.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, **10,** 209–231.

Skehan, P. (1989). *Individual differences in second-language learning*. London: Edward Arnold.

Sorace, A. (1993). Incomplete vs. divergent representations of unaccusativity in non-native grammars of Italian. *Second Language Research*, **9,** 22–47.

Towell, R., & Hawkins, R. (1994). *Approaches to second language acquisition*. Clevedon, UK: Multilingual Matters.

Vainikka, A., & Young-Scholten, M. (1996). The early stages in adult L2 syntax: Additional evidence from Romance speakers. *Second Language Research*, **12,** 140–176.

van Els, T., Bongaerts, T., Extra, G., van Os, C., & Janssen-van Dieten, A.-M. (1984). *Applied linguistics and the learning and teaching of foreign languages*. Baltimore: Edward Arnold.

van Wuijtswinkel, K. (1994). *Critical period effects on the acquisition of grammatical competence in a second language*. Unpublished thesis, Katholieke Universiteit, Nijmegen, The Netherlands.

Weber-Fox, C. M., & Neville, H. J. (1999). Functional neural subsystems are differentially affected by delays in second-language immersion: ERP and behavioral evidence in bilingual speakers. In D. Birdsong (Ed.), *The Critical Period Hypothesis and second language acquisition* (pp. 23–38). Mahwah, NJ: Erlbaum.

Weinert, F. E., & Perner, J. (1996). Cognitive development. In *The lifespan development of individuals: behavioral, neurobiological, and psychosocial perspectives: A synthesis* (pp. 207–222). Cambridge, UK: Cambridge Univ. Press.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, **12,** 238–265.

(Received January 10, 2000)
(Revision received July 5, 2000)