LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Critical Period Claim Revisited: Reanalysis of Hartshorne, Tenenbaum, and Pinker (2018) Suggests Steady Decline and Learner-Type Differences

Frans van der Slik,[a,b] Job Schepens,[c] Theo Bongaerts,[a] and Roeland van Hout[a]

[a]Centre for Language Studies, Radboud University, Nijmegen, [b]Research Focus Area: Understanding and Processing Language in Complex Settings, North-West University, South Africa, and [c]Center for Research on Education and School Development, TU Dortmund University

Abstract: A reanalysis of data drawn by Hartshorne, Tenenbaum, and Pinker (2018) from two-thirds of a million English speakers showed that their overall conclusion of one sharply defined critical age at 17.4 for all language learners is based on artificial results. We show that instead of a discontinuous exponential learning with sigmoidal decay (ELSD) model, a continuous ELSD model had a better fit when applied separately to monolinguals, bilinguals, and early immersion learners. Only for nonimmersion learners and later immersion learners did a discontinuous ELSD model have a better fit, with a critical age of 18.6 and 19.0 years of age, respectively. These age effects can be interpreted as schooling effects. We suggest that personal and societal factors, including differences in living circumstances and socialization, may bring about age-specific

We are grateful to Joshua Hartshorne and Tianhu Chen. We could not have done our analyses without their work and help. We are grateful as well to the anonymous reviewers of *Language Learning*, who helped us tremendously to make our arguments more transparent and precise.

Correspondence concerning this article should be addressed to Frans van der Slik, Centre for Language Studies, Radboud University, Erasmusplein 1, 6537 HT Nijmegen, the Netherlands. E-mail: f.v.d.slik@let.ru.nl

The handling editor for this article was Emma Marsden.

DOI: 10.1111/lang.12470

discontinuity patterns in language learning and in language learning rate. The implication is that they are not the outcome of cognition-driven developmental factors leading to one or more critical periods.

Keywords L2 acquisition; critical period; language acquisition

Introduction

It is now more than half a century since Lenneberg (1967) published his seminal work on the critical period hypothesis. During that period, more than 240 scientific publications have been dedicated to this crucial issue.[1] Although various formulations of this hypothesis have been proposed, differing in many nuances, the central core is the existence of a less or more constrained optimal time window that allows learners to reach nativelike levels of proficiency. Beyond this time window, it is argued, reaching nativelike levels becomes extremely difficult, if not impossible.

A recent study by Hartshorne, Tenenbaum, and Pinker (2018), henceforth HTP, identified one sharply defined critical age of 17.4 years followed by a continuously declining language learning rate thereafter. This conclusion appears impressive, for at least five reasons. First, the critical point of 17.4 years that they identified appeared to pertain across all types of language learning: monolingual or bilingual, immersed or instructed, early or late. Furthermore, the same claim was applied to their finding of a declining learning rate after the critical age: this too was found to be independent of the type of learning. The second reason is the amount of data, drawn from almost two-thirds of a million language users who did the same grammaticality judgment test on English: a sample size that is unprecedented. The third reason is that HTP applied a relatively new, powerful computational algorithm called DEoptim (DE = differential evolution; Mullen, Aridia, Gil, Windover, & Cline, 2011), in order to estimate decay in language learning rates in their exponential learning with sigmoidal decay models (ELSD), both of which are explained further below. The fourth reason is that HTP put forward rate of learning—and not ultimate attainment—as the decisive parameter in determining critical age. Although this approach is not unique (see Krashen, Long, & Scarcella, 1979; Snow & Hoefnagel-Höhle, 1978), it deviates substantially from the bulk of studies on the critical period hypothesis, which use the attainment scores or ratings of native speakers as decisive yardsticks. The fifth reason is that the model selected had a remarkable amount of explained variance of no less than 89%.

Despite our acknowledgement of the positive aspects of HTP, we nevertheless felt it to be necessary to challenge their conclusion that there is one overall sharply delineated critical age in language learning. We have two main objections. The first is that HTP defined a continuous learning rate model (see their figure S2, p. 2; S indicates Supplementary Information), but they did not report any outcomes for this model. The occurrence of a discontinuity is a crucial argument in their positive evaluation of the critical period hypothesis. A proper evaluation includes, in our view, comparing the fit of a continuous model with the fit of a discontinuous model.

Our second objection is that they combined four clearly distinct language learner groups in one comprehensive analysis, without carefully checking whether the relationship between age of acquisition and language learning rate is similar in all these groups. To handle group differences, they used the *E* parameter in their decay model. They called *E* the experience discount factor, with different scale values for the four different speaker groups: (a) monolinguals; (b) bilinguals, with an age of onset of 0 years of age; (c) later immersion learners (henceforth immersion learners), who started learning English from age 1 year onwards; and (d) nonimmersion learners, who learned English at school in a non-English-speaking country. Combining different groups can be advantageous, but not when the explanations for the different findings for each group are distinct. The issue here can be illustrated by Simpson's paradox, about associations or relationships that can be found to be entirely different depending on the level of analysis: a paradox that warns researchers that causal inferences in nonexperimental studies are hazardous when investigating associative patterns in different groups. We wanted to test whether the four groups are different beyond the *E* parameter put forward by HTP. Our expectation was that not all groups would fit a discontinuous model.

It is of interest that HTP's conclusion opposes those of two recent overview articles. Mayberry and Kluender (2018) reviewed research on maturational constraints on the capacity to learn languages. They concluded that the age-of-onset effects are substantially different in first language (L1) acquisition from those in second language (L2) acquisition. Delayed L1 acquisition shows that there is a critical period at very young ages. This critical period is not present or traceable in the HTP data. Mayberry and Kluender (2018) concluded that delayed L1 acquisition is much more pervasive and destructive than whatever critical period in L2 acquisition may exist. In addition, they noted that a critical period occurs at a much later age in L2 acquisition than in L1 learning. Abrahamsson, Hyltenstam, and Bylund (2018) presented an extensive overview of research on age of onset and ultimate attainment in acquiring a L2, in order to investigate the critical period hypothesis for L2 acquisition. They observed discontinuities at two cutoff points: 6−7 and 12−13 years of age at onset of acquisition. Both overview studies mentioned other critical periods and pointed out fundamental differences between learner groups.

Finally, before presenting our own analyses, two additional points need to be raised. The first concerns the validity of a written grammaticality judgment test taken on a computer to measure language proficiency. This point was addressed by HTP, who explicitly discussed the way this grammar quiz, as they called it, was constructed. They referred to the high reliability of this language quiz (Cronbach's $alpha = .86$), but they did not discuss its validity in great detail. They mentioned that the whole quiz could be done within 10 minutes, a time limit that they considered to be important in keeping participants motivated. There was no real time limit on doing the quiz. Their reliance on such a test is not unusual. Plonsky, Marsden, Crowther, Gass, and Spinner (2019) recently found that L2 research in general relies very (and, arguably, too) heavily on acceptability or grammaticality judgment tests, given certain well-founded concerns about their validity. Nevertheless, many relevant papers on the critical period hypothesis apply a grammaticality judgment test as a proficiency measure (e.g., DeKeyser, Alfi-Shabtay, & Ravid, 2010; data reanalyzed by Vanhove, 2013). The grammaticality judgment test as employed by HTP is akin to a metalinguistic awareness test (Loewen, 2009; Vafaee, Suzuki, & Kachisnke, 2017). Although various objections can be raised to the grammaticality judgment test, the main aim of our contribution is to prove that, even if this test is accepted as a valid grammatical proficiency test, HTP's conclusion about an overall critical age is not warranted by their own data.

Second, HTP introduced a new computational model (ELSD) and a parameter optimization algorithm (DEoptim). DEoptim is a powerful optimization procedure to trace optimal model parameters. HTP decided to use aggregated data instead of individual data.[2] The authors used age bins of 3 years for the immersion learners and of 1 year for the other learner groups (see the curves in Figure 2 in our Results section, and HTP's figure 4). The advantage of DEoptim is that complex nonlinear models can be implemented that account for changes in rate of learning. Regression analysis, including mixed regression and breakpoint regression, is an alternative, but in its basic form it lacks sensitivity in tracing critical period or critical age boundaries in language acquisition (see Vanhove, 2013, for the application of a regression approach to trace critical age in L2 acquisition).

The Present Study

The discussion of a critical period appears to be deeply embedded in different perspectives on L2 acquisition, and it impinges on fundamental issues, but we have mainly restricted ourselves in this contribution to a reanalysis of the HTP data. First, we explicitly introduced a continuous model in our analyses, to compare it to HTP's discontinuous model. Second, we analyzed the four learner groups separately. We see this separate analysis of groups as a litmus test. If HTP's conclusion of a critical age of 17.4 years holds generally, this critical age should recur in separate analyses of the four learner groups. HTP presented their finding of a "sharply-defined age boundary" (p. 263) at which the rate of language learning sharply declines (the discontinuity) as evidence for the existence of a critical period overall, but they did not discuss the possibility of a variable critical period in different learner groups.

As the computational and statistical analyses needed to evaluate these issues are not commonly used in research on language acquisition, we first provide an explanatory note on some essential technical and mathematical aspects of these analyses that have driven our study.

Explanatory Note About Computational Models

HTP start by depicting linear models for how language learning ability might change with age, in the first four panels of their figure 1 (p. 264). Model D is the simplest, with a constant, nondecreasing ability plotted against age. As an analogy, consider the simple formula for the outcome "distance travelled" (d), which is velocity multiplied by time ($v \times t$), with a plain outcome when velocity is constant. Similarly, if we would take acquired proficiency g (grammatical proficiency), we might define this as *rate* (learning ability) $\times$ t (time), with rate being constant. However, rate (or velocity) itself may change smoothly and linearly, as visualized in Panel A of their figure, or abruptly, as visualized in Panels B and C. The most important feature of those figures is that they distinguish rate (or velocity) as a fundamental, interpretable property. In Panels E–H in figure 1, HTP then present graphs in terms of ultimate attainment (y-axis) plotted against age of first exposure (x-axis); these graphs resemble those seen in other publications (see Birdsong, 2006, and Vanhove, 2013, both presenting panels with ultimate attainment on the y-axis). How can we calculate changing rates in learning and their impact on the outcome of g? When time is the years between current age and age of first exposure, we need to integrate (a mathematical operation) the time variable to establish the relationship over time between rate and proficiency (see equation 1 on p. 268 of HTP). We do not explain this mathematical operation in further detail.

How does the intuitive difference between rate and ultimate attainment, outlined above, relate to the ELSD models that were used to model growth and decay in rates of learning applied by HTP? The common growth models are exponential, a property known to a wider audience nowadays through the figures on the spread of COVID-19. When rate increases, growth curves accelerate. The opposite effect is established by a shrinking rate. Growth models can have different mathematical forms depending on the type of growth pattern they aim to cover. The logistic model is the best-known mathematical form of the so-called exponential sigmoid(al) functions (algebraic functions that give "*S*-shaped" curves with one inflection point). Many growth processes exhibit changing rates of exponential growth, until they reach an inflection point, where the rate of growth decelerates. Examples include population growth, human learning, and language change. A linguistic example of *S*-shaped growth curve analysis is provided by van Veen, Evers-Vermeul, Sanders, and van den Bergh (2009), who investigated the development of causal connectives in L1 acquisition. Decay models are the inverse of growth models. The decay model of procedural memory potential presented by Abrahamsson et al. (2018, figure 1.5, p. 42), illustrating age differences in language learning, is strikingly similar to the continuous models we used in the current study.

What complications did HTP have to face in applying ELSD models? They had data about proficiency level (*g*), the start of the learning process (time of exposure), and the current age of the participants, but they needed a mathematical model to estimate a series of unobserved, hidden parameters (HTP, p. 268: learning rate, differences between learner groups [experience discount factor], shape parameters of the sigmoid function). In addition, they wanted to determine an age boundary, the critical age, at which the rate shifts (expressed by two different mathematical functions, one applying before and the other applying after the critical age; see equation 2 on p. 268 of HTP). Straightforward mathematical estimations of the underlying parameters are impossible in this situation. They therefore used a powerful computational algorithm to estimate the parameters they needed. This algorithm, called differential evolution (DE), optimizes predefined mathematical functions (see, for example, Rodriguez-Mier, 2017; for an application of this algorithm to biological growth, see Cao, Shi, Li, & Chen, 2019). It is a successful algorithm for solving nonlinear optimization problems in computational intelligence research, but also in the domain of machine and deep learning. After a random start, the algorithm proceeds step by step (iteration after iteration) to find a new and better solution until an optimal fit is found. This algorithm is powerful in finding optimal solutions for complex functions with continuous parameters, such as age and language proficiency (the variable *g*).

Crucial in finding the optimal solutions in the models of HTP is the concept of ultimate attainment in proficiency. The level of proficiency is considered as “below native level” (where native level is the maximum score on the outcome measure) if the rate of learning becomes zero before the native level is reached. The level reached is treated as the ultimate attainment for that particular group of learners. Figure 2 in our Results section (see also figure 4 in HTP, p. 268) shows lower levels of ultimate attainment in different learner groups. The attainment is measured in terms of accuracy on the grammaticality judgment test that returns an outcome between 0 and 1, in fact a proportion of correct answers (*g*). The figures on accuracy give the logit (or log-odds) of this proportion: $\log(p/[1-p])$ (HTP, p. 267). This transformation is typical of logistic regression. It is nowadays the standard way of dealing with proportions in statistical analysis. A crucial advantage of the logit is that its value may range from minus to positive infinity.

Method

Data

It is becoming common practice that scientific journals urge researchers to publish their data, allowing for replication research and reanalysis (Marsden, Morgan-Short, Thompson, & Abugaber, 2018). The HTP paper meets this new standard. Their data are publicly available at https://osf.io/pyb8s/. The data were collected by means of a grammaticality judgment test, which the authors called a grammar quiz, available at http://archive.gameswithwords.org/WhichEnglish/. The R codes used in the present study (van der Slik, Schepens, Bongaerts, & van Hout, 2021) are freely available at both IRIS (http://www.iris-database.org) and the Open Science Framework (https://osf.io/gqm87/).

Participants

For the definitions of the four groups of language learners selected by HTP, we refer readers to their paper and its Supplementary Information. They can be briefly described as follows (note that two of the groups, early bilinguals and later immersion learners, were combined by HTP as “immersion learners,” whereas we made a different distinction, i.e., between bilinguals and immersion learners):

- Monolinguals ($n = 244{,}840$), who grew up speaking English only, age of first exposure (AoE) = 0;
- HTP’s “immersion learners” ($n = 44{,}412$), including those who acquired English and another language from birth onwards (AoE = 0; bilinguals in our terminology), and later learners who spent at least 90% of their life in an English-speaking country, learning English in that setting from AoE ≥ 1 (immersion learners in our terminology [3]); and

- Nonimmersion learners ($n = 257{,}998$), who spent at most 10% of their life in an English-speaking country, learning English in a non-English speaking country from AoE ≥ 4.

More than 100,000 English speakers of the total sample could not be classified as belonging to one of those groups (see Table 2 in our Results section); thus HTP included in their analysis 547,250 speakers of the two-thirds-of-a-million sample mentioned in the title of their article.

Measures

Grammaticality Judgment Test

HTP took what they called a "shotgun approach" (p. 267) to assess a variety of syntactic phenomena such as tense, word order, and determiners (see HTP, p. 267, for a comprehensive description). In total, 35 questions were used, of which 27 in fact comprised 124 distinct questions, whereas the remaining 8 were in dichotomous format. Of these 132 items, 95 were actually selected for analysis, based on the criterion that at least 70% of the native English-speaking adults gave the same response. A full description of these items and their coding can be found in the Supplementary Information of the HTP paper. The test has a high reliability (Cronbach's alpha = .86). The resulting scores (proportion correct answers, ranging between 0 and 1), were transformed into a logit value (log-odds) (*Mean* = 3.05; *SD* = 1.02).

Age-related measures

In the analyses, three measures or variables are distinguished in relation to the age of the participants:

- Age of first exposure (AoE), ranging between 0−30 years of age,
- Current age, ranging between 7−70 years of age,
- Time of exposure, the difference between the two previous variables: current age – AoE.

Results

We present the results of our reanalyses of the HTP data in two parts. In the first part we present the outcomes of our replication of the best-fitting discontinuous ELSD model as applied by HTP (see their figure 4, p. 268, and their figure S6, p. 8) and the outcomes of the best-fitting continuous ELSD model we found for all four learner groups jointly. In the second part we reanalyze the data for the four learner groups separately, again comparing the outcomes of discontinuous and continuous ELSD models, and then further split the immersion learners into an early and a later group. In their equation 2 (p. 268), HTP present two functions to be optimized: (a) a constant rate from birth to a critical age (t_c) somewhere on the age line, followed by (b) a sigmoidal changing rate model with rate decelerating to 0. The difference in rate between the estimates of the two functions may show a sudden drop. This gives t_c two different properties: the switch from a constant rate to a decelerating rate plus the difference in rate at t_c (see figure 4 in HTP, p. 268). In a continuous model, only the second function is activated for the whole timeline. The sigmoid has two shape parameters, α (steepness) and δ (center sigmoid).

Part 1: All Four Learner Groups Analyzed Together

The first analysis we present is a replication of the discontinuous model that HTP found to be the best one plus the best continuous model we found when analyzing all four groups together. The model parameters are given in Table 1. The Akaike information criterion (AIC; Akaike, 1974) is the most relevant measure in evaluating the model in question; the lower its value, the better. We use the *relative likelihood*[4] to evaluate the model fit of the continuous ELSD model as compared to the discontinuous ELSD model. The research hypothesis of a sharp age boundary (t_c) is represented by the discontinuous model; the null hypothesis of no sharp age boundary (t_c) is represented by the continuous model. The discontinuous model appears to outperform the continuous model in terms of relative likelihood (the discontinuous model has an extremely high likelihood value in comparison to the continuous model). The changing rates of the two models are visualized in Figure 1.

Table 1 Outcomes of the best-fitting exponential learning with sigmoidal decay (ELSD) discontinuous model and the best-fitting ELSD continuous model for all four learner groups combined

ELSD model	r	α	δ	E_{sb}	E_{li}	E_{ni}	t_c	R^2	AIC	Rel. LL of D model
Discontinuous	.19	.09	−0.44	.63	1.00	.29	18.0	.92	−6,149	$> 10^5$
Continuous	.16	.24	30.80	.70	.80	.24	1.0	.87	−5,406	

Note: Number of degrees of freedom of discontinuous model = number of degrees of freedom of continuous model + 2. The parameters α and δ represent steepness and center of the sigmoid function. The three E parameters refer to three groups: sb = bilinguals, li = immersion learners, ni = nonimmersion learners; the E value of the monolinguals is fixed at 1.00. The parameter t_c is the age boundary (critical age). AIC = Akaike information criterion; rel. LL of D model = relative likelihood.

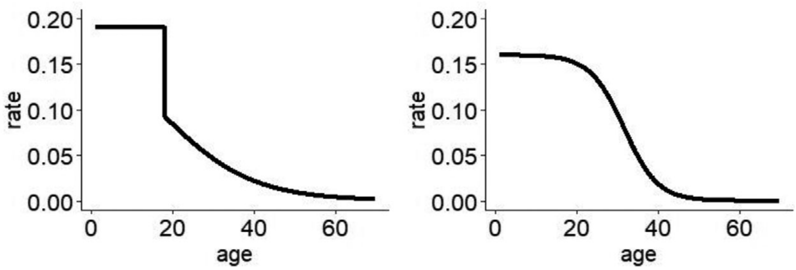

Figure 1 The learning rates of the best-fitting discontinuous exponential learning with sigmoidal decay (ELSD) model (left) and the best-fitting continuous ELSD model (right) for all four learner groups combined.

Before comparing the two models in more detail, we will first discuss the small differences between the discontinuous model used by HTP and our discontinuous model. The discontinuous model we obtained actually has a slightly higher R^2 value of .92 (HTP found .89). The critical age (t_c) is 18.0 years in our model and 17.4 in the HTP model. When the number of iterations is 500, as performed by HTP, the critical age is not reproducible without knowing the seed (an arbitrary value needed to start the optimization; HTP did not report this value), and it can vary rather substantially. We did find the critical age to become stable and reproducible, independent from the factual seed value used, after 2,500 iterations, with a critical age of 18.0.

The parameter t_c (age boundary) in the continuous model is set at 1, to force a continuous model. The continuous model is a regular sigmoid function, showing a decreasing rate of learning, starting at .16 and approaching 0 after 50 years of age, after an accelerating decrease between 18 and 40 years of age. The discontinuous model has three components:[5] (a) a constant rate of learning of .19, up to (b) a critical age at 18.0 years of age with a sharp drop halving the learning rate, after which (c) a new decreasing rate function gradually becomes 0 at about the age of 60 years.

Although mathematically the discontinuous (= D) model seems to be the best, there are logical reasons to prefer the continuous (= C) model. The continuous function is easier to interpret than a sudden drop at 18.0 years. It indicates a process distributed over a longer period, and it shows at the same time that language learning is much faster at a younger age, slowing down when the learner gets older. It also means that a sharply delineated critical age is not necessarily required to cover the data and that a decrease in learning rate can be described as a regularly decelerating process extended over a long time period.

The initial rates (.19 and .16, for the D and C models, respectively) are the initial rates for the monolinguals, since they have an *E* value of 1. The initial rates for the other groups can be calculated by multiplying the monolinguals' rates by the values of the *E* parameter for each of the groups. In Model D, both monolinguals and immersion learners have an *E* value of 1. The discount factor is multiplied by the rate to compute the actual rate of learning when the input or experience is less than 1. That seems to make sense for the nonimmersion learners (.29) and for the bilinguals (.63), because they both have to learn two languages, which are in competition. The same value of 1 for both monolinguals and immersion learners is counterintuitive. Why should immersion learners have a higher rate of learning than bilinguals? This may indicate a model misspecification. The C model has the logical outcome that all three learner groups have a value lower than 1.

Because the mathematical analysis and logical analysis of the C and D models are contradictory and thus inconclusive, it makes sense to see how well the two models apply to learner groups when analyzed separately. This can tell us how different the learner groups are in actual fact.

There are two more technical arguments for being hesitant in accepting a comprehensive model applying to all sorts of learners. First, the data points analyzed (as also done by HTP) are data aggregated by cross-tabulating current age by starting age of language learning within the four groups that are distinguished. There are 547,288 learners, whose results are aggregated in 1,696 data points. The advantage is that aggregation produces cleaner and more clear-cut patterns, leading to much higher R^2 values on the aggregated level than on the individual speaker level. We should not overestimate the meaning of the artificially high values for R^2 in aggregated models.

Second, the number of aggregated data points has an uneven distribution. All monolinguals and bilinguals start at the age of 0. The data points for these groups, 64, refer to current age. The nonimmersion learners have many more data points than the three other learner groups, as can be seen in Table 2. The consequence is that the nonimmersion learners, being by far the largest group in number of data points, mainly determine the age boundary found by HTP and by us in analyzing all learner groups with one comprehensive model.

Part 2: All Four Learner Groups Analyzed Separately

Turning to the next analysis, we present the outcomes for the best-fitting continuous (C) and discontinuous (D) models for the four separate language learner groups. The *E* parameter was fixed at 1, to neutralize and exclude this

Table 2 The number of learners and the number of data points in the four language learner groups

Language learner group	*n* of learners in current study (in HTP)		*k* of data points
Monolinguals	244,840	(244,840)	64
Bilinguals	30,347	(unspecified)	64
Immersion learners	14,099	(unspecified)	398
Nonimmersion learners	258,002	(257,998)	1,170
Total	547,288	(547,250)	1,696

Note: Our total number of learners is 38 higher than the number (given in parentheses) reported by Hartshorne et al. (2018; abbreviated as HTP). The reason for this difference is unclear.

Table 3 Outcomes of the best-fitting discontinuous exponential learning with sigmoidal decay (ELSD) model and the best-fitting continuous ELSD model in four groups of speakers analyzed separately

Group	Model	r	α	δ	t_c	R^2	AIC	Rel. LL of D model
Immersion learners	Discontinuous	.15	.08	8.92	15.6	.92	−1,567	$> 10^5$
	Continuous	.11	.17	34.25	1.0	.89	−1,427	
Nonimmersion learners	Discontinuous	.06	.10	−3.34	18.6	.85	−4,368	$> 10^5$
	Continuous	.04	.32	27.75	1.0	.74	−3,754	
Monolinguals	Discontinuous	.15	1.0	50.00	40.0	.73	−262	0.16
	Continuous	.15	1.0	50.00	0.0	.73	−266	
Bilinguals	Discontinuous	.10	1.0	−0.00	38.0	.92	−314	0.14
	Continuous	.10	1.0	38.83	0.0	.92	−318	

Note: Number of degrees of freedom of discontinuous model = number of degrees of freedom of continuous model + 2. The parameters α and δ represent steepness and center of the sigmoid function. The parameter t_c is the age boundary (critical age). AIC = Akaike information criterion; rel. LL of D model = relative likelihood.

parameter, as we do not compare groups in this analysis. The other model parameters are given in Table 3.

Table 3 shows that, for the data from the immersion and nonimmersion learner groups, a discontinuous model performs better than a continuous model, because for each group the AIC is lower and the relative likelihood is well above 1. The discontinuous model is superior for the nonimmersion learners with an age boundary of 18.6 years and for the immersion learners with an age boundary of 15.6. The latter age boundary shows a difference of almost 2 years from the outcome reported by HTP (17.4 years). There is no straightforward statistical way to test this age difference. They belong to partly overlapping learner groups (immersion learners vs. all groups combined).

The continuous models of the monolinguals and bilinguals perform better than the discontinuous models, because for each group the AIC is lower and the relative likelihood is below 1. Moreover, the discontinuous models have very high age boundaries (40.0 years for monolinguals and 38.0 years for bilinguals). The outcome of 40 as a critical age for the monolinguals has to be ascribed to the age limit constraint of 40 years of experience implemented in the models (HTP, Supplementary Information, p. 3). Monolinguals have a higher learning rate than bilinguals, which, according to HTP (p. 268), reflects the fact that bilinguals "may receive less English input than monolinguals."

As visualizations of the outcomes, we present the learning rate curves of the discontinuous models of the immersion learners and the nonimmersion learners, the two groups with the larger number of data points, in Figure 2. We also present the raw accuracy figures (performance on the grammaticality judgment test) and the accuracy curves predicted by the model.

The distinction between the two rate patterns is considerable in Figure 2, and the same applies to the ultimate attainment of both language learner groups (expressed as judgment accuracy in Figure 2). The ultimate attainment scores of the immersion group are much higher, and their learning rate pattern is much steeper. Importantly, the rate drop in both models at the age boundary is much less steep than in the overall model (less than .05 here in Figure 2, left panel, and about .10 in the overall model, see Figure 1, left panel). The age boundary is visible in the raw data of the nonimmersion learners in the break of the steep slopes of the orange and some green curves followed by flattened patterns, but such a break is not clearly observable in the raw data of the immersion learners.

Is the critical age of the immersion learners a hard and stable boundary? Figure 2 shows a sharp distinction between early (AoE: 1−9 years) and later (AoE: 10−30 years) immersion learners. Additional data show that more than 60% of immersion learners who started acquiring English before their teens (AoE: 1−9 years) reported using English as their sole medium of communication, whereas these proportions decline for learners who started learning English at later ages (less than 40% for those who started learning English after AoE: 20 years). When the critical age is a robust boundary, it should recur after splitting the immersion group of learners into these two subcategories. Thus, we repeated our analyses for an early (AoE: 1−9 years; $n = 11{,}594$) and a later (AoE: 10−30 years; $n = 2{,}505$) group of immersion learners. The results are summarized in Table 4.

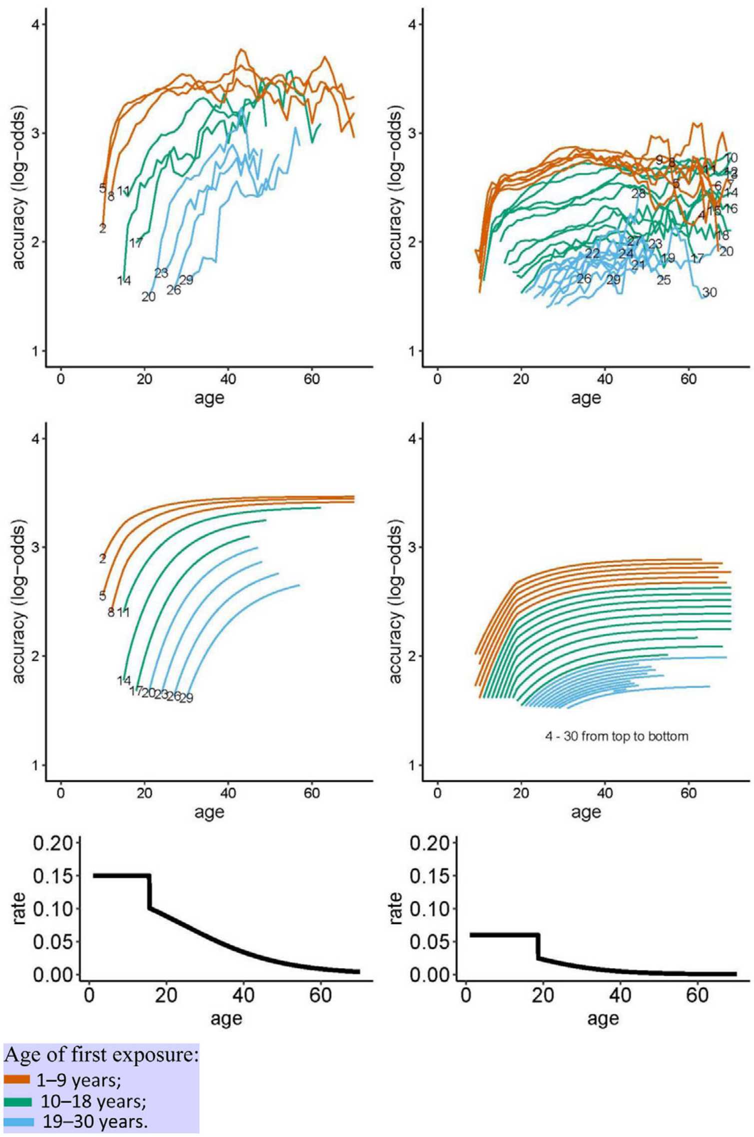

Figure 2 Accuracy curves, raw (top) and predicted (middle), and learning rates (bottom) of the best-fitting discontinuous exponential learning with sigmoidal decay (ELSD) model for the immersion learners (left) and the best-fitting discontinuous ELSD model for the nonimmersion learners (right). [Color figure can be viewed at wileyonlinelibrary.com]

Table 4 Outcomes of the best-fitting discontinuous exponential learning with sigmoidal decay (ELSD) model and the best-fitting continuous ELSD model for early and later immersion learners

Group	Model	r	α	δ	t_c	R^2	AIC	Rel. LL of D model
Early immersion learners	Discontinuous	.15	.93	−38.98	27.1	.65	−712	0.15
(AoE: 1−9)	Continuous	.15	1.00	26.10	1.0	.65	−715	
Later immersion learners	Discontinuous	.12	.08	10.44	19.0	.92	−881	$> 10^5$
(AoE: 10−30)	Continuous	.09	.17	36.16	1.0	.90	−832	

Note: Number of degrees of freedom of discontinuous model = number of degrees of freedom of continuous model + 2. The parameters α and δ represent steepness and center of the sigmoid function. The parameter t_c is the age boundary (critical age). AIC = Akaike information criterion; rel. LL of D model = relative likelihood; AoE = age of first exposure. For the early immersion learners, the parameters α and δ of the discontinuous model are unstable. The remaining parameters are stable.

Table 4 suggests a redefinition of the immersion group. The early learners match the bilingual learners, having the same rate of learning, whereas the later learners fit the nonimmersion group, with a higher learning rate but with about the same age boundary (19.0 years of age instead of 18.6). We analyzed all age bins (2, 5, 8, … 29) separately, and the results confirm that the turning point between a continuous and discontinuous model is indeed between 7−9 and 10−12 years old. The immersion group appears to be heterogeneous with respect to the age boundary, and thus we need to make a distinction between early and later immersion learners. The best model for the early immersion learners is continuous, whereas the best model for the later immersion learners is discontinuous. The two models are visualized in Figure 3.

The early immersion learners now share their continuous model with the monolingual and bilingual learners. The later immersion learners share their discontinuous model with the nonimmersion learners, with a similar age boundary of 19.0 years (18.6 years for the nonimmersion learners), which is substantially different from the age boundary of 15.6 years in the overall model of the immersion learners.

Discussion

HTP identified a sharply defined boundary age in grammatical learning and a steady decline thereafter, for all groups of language learners: an astonishing outcome. This outcome is seen as evidence for a critical period for language

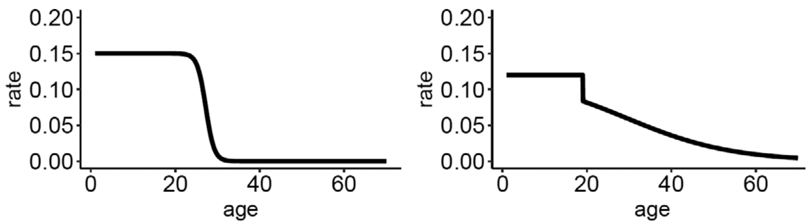

Figure 3 Curves of the learning rate of the best-fitting continuous exponential learning with sigmoidal decay (ELSD) model for the early immersion learners (age of first exposure: 1−9; left) and of the best-fitting discontinuous ELSD model for the later immersion learners (age of first exposure: 10−30; right).

learning in general. They stated that discussions to date concerning mother tongue acquisition, early and late L2 acquisition, monolinguals and bilinguals, learning in school, immersion learning, and the many proficiency differences between the learners involved had been misguided in focusing on age of onset and on ultimate attainment instead of on rate of language learning. HTP convincingly argued that rate of learning was a missing link. Mathematical learning or growth models are defined not only by their end state but also by their rate. An important issue is how to interpret rate. Technically, it is plainly a parameter that may be constant or variable. From a language learning point of view, one could argue, as HTP in fact did, that it should be reconceptualized as learning capacity that is of course sensitive to language input or learning experience, incorporated in the discount parameter *E* of HTP. This discount factor, however, is not sufficiently powerful to trace different learning rate models among the four language learner groups (see, for example, the overview studies of Abrahamsson et al., 2018, and Mayberry & Kluender, 2018, which stress the fundamental differences between language learner groups). HTP's reported outcome—a general age boundary drop at 17.4 (18.0 in our analysis), independent of the kind of language learning, with a sudden rate drop that halves the rate of learning—is implausible.

Despite HTP's interpretation of this age boundary being applicable to all groups, including monolinguals and bilinguals, in fact no theories posit a sharply delimited critical period for normally developing monolinguals and bilinguals. Theories of L1 acquisition emphasize that L1 learning is a very fast process, but the HTP data cannot give direct answers on early acquisition. Most of the participants doing the written grammaticality judgment test were 10 years of age or older, with the youngest participant aged 7. On the basis of our separate analyses of the four groups, we conclude that a continuous model can better explain the outcomes for the monolinguals and bilinguals. As we noted in the Introduction, the grammaticality judgment test employed in the HTP paper is a kind of metalinguistic awareness test: the participants are asked to reflect about language and their own language use (Loewen, 2009; Vafaee et al., 2017). Obviously, we need data sets on various types of language proficiency skills in order to test in great precision and with sufficient power the course of language learning rates across all ages and its relationship to ultimate attainment.

Nevertheless, we found relevant and sometimes even remarkable results. The nonimmersion learners are incontestably better modeled by a discontinuity at 18.6 years of age, again beyond the younger critical periods signaled earlier. This discontinuity (a drop of .05) is much smaller and therefore easier to explain than the discontinuity found by HTP (a drop of .10). We want to suggest a societal rather than a biological explanation, a possibility also suggested, though not foregrounded, by HTP (p. 275: critical period as "an epiphenomenon of culture"). This critical age marks the end of secondary schooling in most countries, at least developed countries, from where most of the data came. In many countries, English is the only or the most important L2 in primary and secondary education. Learning English later in life, in a non-English environment, does not offer the instructional settings of regular schooling, thereby decreasing the opportunities for learning English at a higher level (at least, for the aspects of English knowledge that may have been elicited by the judgment test used). The most obvious interpretation of the outcomes for nonimmersion learners is that 18.6 marks a schooling effect (see Flege, 2019, for similar arguments).

We found a discontinuous model for the immersion learners as well. Our discontinuous model returns a critical age of 15 years, which is higher than the critical periods (6−7 and 12−13 years of age at onset of acquisition) mentioned by Abrahamsson et al. (2018). The number of data points for immersion learners is much lower than that for the nonimmersion learners. Age of onset groups were put together in bins of 3 years, with the curve labeled "2" representing the range of 1 to 3 years of age at first exposure, and so on (HTP did the same). Because of the age patterns in the immersion learners, we split these learners into an early group (AoE: 1−9 years) and a later group (AoE: 10−30 years). The early group matches the monolinguals and bilingual learners, being best described by a continuous model.

The later group of immersion learners fits the noncontinuous pattern of the nonimmersion learners, with a comparable age boundary (19.0 vs. 18.6). The age boundary of 15.6 years of age found for the whole group of immersion learners is an artifact of combining two different types of learners: early versus late immersion. This raises the question of how powerful ELSD models are in testing and discriminating among different models. We would like to suggest that including the data of individual learners (rather than aggregated data) in a mixed regression analysis would be more helpful in testing discontinuities; and yet, on the other hand, we expect that this type of analysis would not be powerful enough to uncover the parameters of learning rates and their differences.

The age boundary of the later immersion learners is later than that predicted by existing theories relating to critical ages for L2 acquisition (6−7 and 12−13 years of age; Abrahamsson et al., 2018). For these later immersion learners (just as for the nonimmersion learners), we suggest a societal rather than a biological explanation. This age boundary marks a change in societal and/or educational status, involving different sorts of communicative contexts, with a much higher learning rate for later immersion learners than for nonimmersion learners because of the greater amount of input.

Vanhove (2013, p. 1) concluded that age patterns in L2 acquisition are not governed by a critical period, given the principle of parsimony and the lack of convincing evidence. HTP drew the opposite conclusion, given the evidence from their impressive database. After analyzing the same database in more detail, we draw the same conclusion as Vanhove (2013), at least to the extent that is possible by drawing on the HTP data set. Our conclusion is that a continuous rate model is the learning model underlying grammaticality judgment accuracy in general. The decay model in question might resemble the decay age model of procedural memory potential presented by Abrahamsson et al. (2018), figure 5.1, a sigmoidal exponential decay model) to deal with changing age effects in language learning. In a continuous model, language learning potential or capacity decreases continuously over a long time period, without any age boundary or sharp drop at a specific age. The model might resemble the development of other cognitive capacities, such as the gradual decline over age of the procedural memory potential. Differences between learners in ultimate attainment are the outcome of many factors, including age of onset and amount and quality of input, but also changes in learners' personal environment and in the social context in which they live, learn, and work. One could think, for example, of different upbringing, schooling, or work experiences. Such personal and societal factors may bring about discontinuities in the process of learning and in learning rate. However typical human language is, these learner data on grammatical judgment accuracy certainly do not support a discontinuous model in the biological course of language learning.[6]

We strongly advocate the use of the terminology *continuous* and *discontinuous models* in L2 studies. Such a framework gives us the capacity to discuss a wide range of growth and decay models in detail, including models with a steep decay curve (see Figure 3 for the early immersion learners).

Limitations and Future Directions

HTP used a grammaticality judgment test to measure proficiency. We have already mentioned in the Introduction concerns about the methodological and validity issues of these tests (Plonsky et al., 2019). These concerns remain relevant, but we believed that the best way to reassess HTP's conclusions was to appropriately reanalyze the enormous database used by HTP.

In analyzing the data, we realized that DEoptim provides more options than the ones exploited by HTP; however, the authors deserve praise for applying a growth model to language learning in combination with a powerful optimization algorithm (namely, differential evolution). This shows the benefits of using new insights and tools from other disciplines. The intricacy of the many internal and external factors in language learning should lead L2 researchers to turn more to mathematical and statistical techniques in order to analyze dynamic, nonlinear systems. Dörnyei, MacIntyre, and Henry (2014) have observed the emergence of a dynamic turn in L2 research. We believe that the estimation of changing rates in continuous and discontinuous models demonstrates how to address dynamic, nonlinear issues.

It is necessary to continue analyzing large data sets to address fundamental issues in L2 research. The lack of power given the small sizes of databases gathered by means of experimental research was a central topic for Vanhove (2013) in discussing research on the critical period. He applied regression techniques, directly based on the data of individual learners. Such techniques offer the possibility of including individual learner characteristics in analyses. We have also previously applied mixed-effects regression analysis to individual learner data, including social learner characteristics and their L1s and L2s in learning Dutch as an additional language (Schepens, van der Slik, & van Hout, 2016; Schepens, van Hout, & Jaeger, 2020; van der Slik, 2010; van der Slik, van Hout, & Schepens, 2015, 2019). In contrast, HTP aggregated data in age bins and analyzed individual learner characteristics in other ways (regression, permutation tests). To be sure that HTP's claim about one clearly defined age boundary at which a sharper rate of decline begins, is credible, a future study should test the decay model on the individual level. HTP argued that analyzing on the level of individual learners would not change their conclusion, without actually testing this conclusion in an overall model. We prefer to conclude that we should proceed by developing and applying additional analysis techniques to explore the full richness of such large data sets.

Analyzing big data is, of course, not a panacea for validity problems. First, Internet-based gathering of data typically lacks the control mechanisms that laboratory or classroom experiments usually have. Internet test takers are self-selected, meaning that interested language learners are overrepresented (Flege, 2019). As a consequence, the results may not reflect the L2 acquisition process of the average English language learner. Second, Internet test takers do not answer, or do not properly answer, all questions, particularly not in a free task. More than 100,000 language learners in the HTP database could not be classified as belonging to one of the four groups because key information was missing. One might speculate that these uncategorized learners were less motivated or less proficient learners than the categorized ones.

Next, the grammaticality judgment test used by HTP can be classified as an untimed test (Godfroid et al., 2015), because there is no limit on the time the test taker is granted to do the test. This may affect the results of the test in different ways depending on which group the language learner belongs to (see Godfroid et al., 2015). HTP collected detailed information on when test takers started to take the test. They also might have information on when the test takers finished the test, though this is not mentioned on the OSF site with the data. If the data are available, it would be worthwhile to check whether test duration has an effect on the results.

A final comment is on HTP's decision to make use of aggregated data at the age level instead of using individual data. (This issue is different to that of aggregating the data of the four learner groups in one analysis.) The authors argued that aggregation on age was necessary because otherwise the outcomes would be affected by the huge number of monolinguals. This might be true, but aggregating the data meant that the outcomes of the nonimmersion learners unduly influenced the outcomes, as we have demonstrated here. This is not to say that we reject aggregation categorically as a fruitful way of analyzing data. One positive consequence of aggregating might be that the outcomes are less affected by sample-specific distributions of the data. Another positive consequence of aggregating is that the estimation process is far less time-consuming. However, bearing Simpson's paradox in mind, researchers should avoid aggregating groups that are qualitatively different. In the present article we have tried to follow this principle.

Conclusion

Hartshorne et al. (2018) identified a sharply defined boundary age (or critical period) in grammatical learning after aggregating data from four different learner groups: monolinguals, bilinguals, immersion learners, and nonimmersion learners. Their resulting learning model was discontinuous, reflecting a critical age period in language learning across all learner types. This discontinuity supposedly marks a sudden drop in rate of learning, a concept that they successfully implemented in their learning model.

We have showed here that including four learner groups in the same analysis returns artificial results. We analyzed the four groups separately, providing evidence that a continuous model better fits the learning patterns of monolinguals, bilinguals, and learners who became immersed at a young age (under 10 years old). Discontinuities remained in the analysis of the data from learners who became immersed when they were older and the nonimmersion learners. We suggest that personal and societal factors, including differences in living circumstances and socialization, may bring about age-specific discontinuity patterns in language learning and in language learning rates. A key implication of our analyses and arguments is that these discontinuity patterns are not the outcome of cognition-driven developmental factors leading to one or more critical periods.[7]

Final revised version accepted 8 June 2021

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at http://www.iris-database.org and https://osf.io/pyb8s/. All proprietary materials have been precisely identified in the manuscript.

Notes

1 Google Scholar (July 15, 2021). Search term: (“critical period hypothesis” or “sensitive period hypothesis”) and (“second language”).

2 Their main argument for doing so seems to be to reduce the impact of the monolinguals who are overwhelmingly present in the test results. Their decision to aggregate all groups of learners is questionable; we address this issue in the Limitations and Future Directions section.

3 Further on, we are going to make a distinction between early (age of first experience below 10 years of age) and later immersion learners (age of first experience 10 years or older).
4 The relative likelihood is defined as $\exp([AIC_C - AIC_D]/2)$, where AIC_C stands for the AIC of the continuous ELSD model, representing the null hypothesis, and AIC_D stands for the AIC of the discontinuous ELSD model, representing the research hypothesis. A value higher than 1 indicates a higher likelihood of the discontinuous model (D); a value lower than 1 indicates a higher likelihood of the continuous model (C; see Burnham & Anderson, 2002). The formula is as follows: $AIC = n \times \log(\text{residual sum of squares}/n) + 2 \times k$.
5 The discontinuous model needs to fit three components, the continuous model only one (cf. Figure 1). That explains the difference of two degrees of freedom.
6 There are numerous studies that advocate nonbiological explanations for differences in ultimate attainment (see, for example, Bongaerts, van Summeren, Planken, & Schils, 1997).
7 After the acceptance of our article by *Language Learning*, Chen and Hartshorne (2021) published their analysis of an expanded data set. As suggested by Frank (2018) they applied an item response theory (IRT) analysis of the HTP grammaticality judgment data (instead of the classical reliability analysis), they calculated interval estimations of rate changes, and they applied a segmented discontinuous ELSD model instead of HTP's discontinuous model. These additions do not substantially change the result: a critical age of 18.4 years for all language learners. More importantly, again they did not test this critical age for the learner groups separately, and, again, they did not test if the application of their segmented model has resulted in a significant improvement in model fit as compared to the continuous model or even the original HTP discontinuous model. It is therefore difficult to see what the added value is of the Chen and Hartshorne (2021) study.

References

Abrahamsson, N., Hyltenstam, K., & Bylund, E. (2018). Age effects on language acquisition, retention and loss: Key hypotheses and findings. In K. Hyltenstam, I. Bartning, & L. Fant (Eds.), *High-level language proficiency in second language and multilingual contexts* (pp. 16–49). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781316809686.002

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*, 9–49. https://doi.org/10.1111/j.1467-9922.2006.00353.x

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*, 447–465. https://doi.org/10.1017/S0272263197004026

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretical approach* (2nd ed.). New York, NY: Springer-Verlag.

Cao, L., Shi, P., Li, L., & Chen, G. (2019). A new flexible sigmoidal growth model. *Symmetry*, *11*, 204. https://doi.org/10.3390/sym11020204

Chen, T., & Hartshorne, J. K. (2021). More evidence from over 1.1 million subjects that the critical period for syntax closes in late adolescence. *Cognition*, *214*, 104706. https://doi.org/10.1016/j.cognition.2021.104706

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguist*, *31*, 413–438. https://doi.org/10.1017/S0142716410000056

Dörnyei, Z., MacIntyre, P. D., & Henry, A. (2014). Introduction: Applying complex dynamic systems principles to empirical research on L2 motivation. In Z. Dörnyei, P. D. MacIntyre, & A. Alastair Henry (Eds.), *Motivational dynamics in language learning* (pp. 1–7). Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781783092574-003

Flege, J. E. (2019). A non-critical period for second-language learning. In A. M. Nyvad, M. Hejná, A. Højen, A. Bothe Jespersen, & M. Hjortshøj Sørensen (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 501–541). Aarhus, Denmark: Dept of English, School of Communication & Culture, Aarhus University.

Frank, M. C. (2018). With great data comes great (theoretical) opportunity. *Trends in Cognitive Sciences*, *22*, 669–671. https://doi.org/10.1016/j.tics.2018.05.011

Godfroid, A., Loewen, S., Jung, S., Park, J-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgment measure distinct types of knowledge. *Studies in Second Language Acquisition*, *37*(special issue 2), 269–297. https://doi.org/10.1017/S0272263114000850

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Krashen, S. D., Long, M. H., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly*, *13*, 573–582. https://doi.org/10.2307/3586451

Lenneberg, E. (1967). *Biological foundations of language*. New York, NY: Wiley.

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94–112). Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781847691767-006

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, *68*, 321–391. https://doi.org/10.1111/lang.12286

Mayberry, R. I., & Kluender, R. (2018). Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition*, *21*, 886–905. https://doi.org/10.1017/S1366728917000724

Mullen, K., Aridia, D., Gil, D., Windover, D., & Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, *40*, 1–26. https://doi.org/10.18637/jss.v040.i06

Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2019). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, *36*, 583–621. https://doi.org/10.1177/0267658319828413

Rodriguez-Mier, P. (2017). A tutorial on differential evolution with Python [Blog post]. Retrieved from https://pablormier.github.io/2017/09/05/a-tutorial-on-differential-evolution-with-python/

Schepens, J., van der Slik, F., & van Hout, R. (2016). L1 and L2 distance effects in learning L3 Dutch. *Language Learning*, *66*, 224–256. https://doi.org/10.1111/lang.12150

Schepens, J., van Hout, R., & Jaeger, T. F. (2020). Big data suggest strong constraints of linguistic similarity on adult language learning. *Cognition*, *194*, 104056. https://doi.org/10.1016/j.cognition.2019.104056

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, *49*, 1114–1128. https://doi.org/10.2307/1128751

Vafaee, P., Suzuki, Y., & Kachisnke, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition*, *39*, 59–95. https://doi.org/10.1017/S0272263115000455

van der Slik, F. (2010). Acquisition of Dutch as a second language: The explanative power of cognate and genetic linguistic distance measures for 11 West-European first languages. *Studies in Second Language Acquisition*, *32*, 401–432. https://doi.org/10.1017/S0272263110000021

van der Slik, F., Schepens, J., Bongaerts, T., & van Hout, R. (2021). *R code. Materials from "Critical period claim revisited: Reanalysis of Hartshorne, Tenenbaum, and Pinker (2018) suggests steady decline and learner-type differences"* [Software/Analysis code]. IRIS Database, University of York, UK. https://doi.org/10.48316/f126-db62

van der Slik, F., van Hout, R., & Schepens, J. (2015). The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PLoS ONE*, *10*(11), e0142056. https://doi.org/10.1371/journal.pone.0142056 van der Slik, F., van Hout, R., & Schepens, J. (2019). The role of morphological complexity in predicting the learnability of an additional language: The case of La (additional language) Dutch. *Second Language Research*, *35*, 47–70. https://doi.org/10.1177/0267658317691322

Vanhove, J. (2013). The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PLoS ONE*, *8*(7), e69172. https://doi.org/10.1371/journal.pone.0069172

van Veen, R., Evers-Vermeul, J., Sanders, T., & van den Bergh, H. (2009). Parental input and connective acquisition: A growth curve analysis. *First Language*, *29*, 266–288. https://doi.org/10.1177/0142723708101679

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Different Learner Groups Have Different Continuous Rates of Learning English: There Is No Critical Age Period

What This Research Was About and Why It Is Important

Ever since Lenneberg's (1967) pivotal study there has been an ongoing debate on the existence of a critical age in second language learning. Central to this notion is the presence of a less or more constrained time window that allows learners to reach nativelike levels of proficiency. Beyond this time window, reaching nativelike proficiency levels becomes no longer possible. Hartshorne, Tenenbaum, and Pinker (2018), henceforth HTP, analyzed data of more than half a million language learners of English who did an English grammaticality judgment test over the Internet. HTP used these test outcomes as language proficiency scores. They arrived at the conclusion that such a critical age indeed exists and that it is 17.4 years of age and operates independently of the four learner-type groups that participated in their study (i.e., monolinguals, bilinguals, immersion learners, and nonimmersion learners). This claim was refuted by the outcomes of our own reanalysis of the original HTP data. We applied the same statistical approach, but we systematically separated the four learner groups. These results suggest that personal and societal factors, including differences in living circumstances and socialization, might bring about age-related discontinuity patterns in language learning and in language learning rates. Such disruptions are not the outcome of abrupt changes in cognition-driven developmental factors leading to one or more critical periods.

What the Researchers Did

- The original HTP data were reanalyzed, by splitting the data for the four learner groups and analyzing these groups separately.
- Continuous models (statistical models assuming no critical age) of language acquisition were tested to compare their performance to the discontinuous models tested by HTP.

What the Researchers Found

- No critical age was found for monolinguals, bilinguals, and early immersion learners.
- A critical age of 18.6 years of age was found for nonimmersion learners, and an age of 19.0 years for later immersion learners. The discontinuities related to these critical age boundaries can be seen as the result of schooling effects.
- There is no evidence for any critical age. Relationships between age and language acquisition are continuous.

Things to Consider

- Aggregating different language learner groups should be done only after careful inspection of differences between those groups.
- Other proficiency scores, in addition to the outcomes of a grammaticality judgment test, should be used.
- Various statistical techniques are required to study developmental patterns in second language acquisition.

Materials, data, open access article: Materials, data, analysis scripts, and the article are publicly available at https://osf.io/pyb8s/ and https://osf.io/gqm87/; analysis scripts are available at https://www.iris-database.org/

How to cite this summary: van der Slik, F., Schepens, J., Bongaerts, T., & van Hout, R. (2021). Different learner groups have different continuous rates of learning English: There is no critical age period. *OASIS Summary* of van der Slik, Schepens, Bongaerts, & van Hout (2022) in *Language Learning*. https://oasis-database.org.