Commentary

New Approaches to Using Census Data to Test the Critical-Period Hypothesis for Second-Language Acquisition

Edward W. Wiley,[1] Ellen Bialystok,[2] and Kenji Hakuta[3]

[1]University of Colorado–Boulder; [2]York University, Toronto, Ontario, Canada; and [3]University of California, Merced

In a previous article (Hakuta, Bialystok, & Wiley, 2003), we tested the critical-period hypothesis for second-language learning using 1990 U.S. census data. Our findings failed to support a critical period for second-language acquisition, demonstrating greater evidence for a steady decline throughout the life span. Stevens (2004) questioned these findings, suggesting they might have resulted in part from the sampling and estimation procedures used. In this Commentary, we report additional analyses addressing each issue raised by Stevens, demonstrating the failure in each case to generate more compelling evidence for a critical period for second-language acquisition.

DETAILS OF SAMPLE SELECTION

Stevens disagreed with our sample selection, arguing that it magnified our conclusion that success in second-language acquisition decreases monotonically with age. At issue here is the wording of the English-ability question, which prohibited identification of "native speakers" of any non-English language who reported speaking only English. Stevens suggested that the sample be augmented to include individuals who reported speaking no other language besides English. Unfortunately, doing so would cause this analysis of second-language acquisition to include not only native speakers who completely stopped speaking their native language, but also individuals monolingual in English to begin with (i.e., those who were never second-language learners). Nevertheless, to address Stevens's concern, we replicated our analyses with samples selected by the question asking for place of birth (as Stevens suggested). Table 1 shows little incremental variance accounted for by critical-period variables (as was the case in our previous analysis based on primary-language samples). Contrary to Stevens's claim, little evidence supports a critical period when English-only speakers are included in the sample.

ASSESSMENT OF RESULTS

Stevens raised a second issue concerning the methods used for testing models of English ability, implying that the way the tests were done led to the conclusions reported: "Their approach allots the bulk of the association to the linear relationship first, and thus there is a ceiling on the amount of additional variance that can be explained by discontinuities" (p. 216). In fact, rather than being suspect, this process of "allotting" (partitioning) variance is the very process by which significance is established. When considering several possible predictive relationships, one first takes a descriptive look at the data (e.g., Figs. 2 and 3 of our original report) to determine which effects appear strongest (in this case, an age-related decline was clearly strongest). The relative magnitude of effects—not researcher preference—dictates the order in which the successive models are tested and significance is established.[1] The goal of this process is parsimony; variables that add little to model fit (as measured by incremental R^2) are excluded from the model. In this case, critical-period variables added little to model fit.

MEASUREMENT ERROR

Stevens also claimed that the existence of a critical period was obscured by the categorical nature of the age and period-of-

Address correspondence to Edward W. Wiley, 249 UCB, University of Colorado–Boulder, Boulder, CO 80309; e-mail: ed.wiley@colorado.edu.

[1]See Neter, Kutner, Nachtsheim, and Wasserman (1996, pp. 474–478) for specific directions for testing regression discontinuities (such as those consistent with a critical period).

TABLE 1

Variance Explained (R^2) by Regressions Using Samples Based on Region of Birth

	Critical point = age 15				Critical point = age 20			
	Central America		China		Central America		China	
Variable	Partial	Total	Partial	Total	Partial	Total	Partial	Total
Intercept	—	—	—	—	—	—	—	—
5–8 years education	—	—	—	—	—	—	—	—
Some high school	—	—	—	—	—	—	—	—
High school graduate	—	—	—	—	—	—	—	—
Some college	—	.2372	—	.3649	—	.2372	—	.3649
Age of immigration	.0541	.2913	.0897	.4546	.0541	.2913	.0897	.4546
Change in slope	.0058	.2971	.0000	.4546	.0062	.2975	.0002	.4548
Change in mean	.0003	.2974	.0001	.4547	.0000	.2975	.0000	.4548

Note. Analyses are based on data from the U.S. Census Public Use Microdata Sample (U.S. Bureau of the Census, 1992). Sample sizes: Central America, n = 2,132,071; China, n = 132,501.

TABLE 2

Variance Explained (R^2) by Regressions Using Samples Based on Language Spoken at Home

	Critical point = age 15				Critical point = age 20			
	Spanish		Chinese		Spanish		Chinese	
Variable	Partial	Total	Partial	Total	Partial	Total	Partial	Total
Intercept	—	—	—	—	—	—	—	—
5–8 years education	—	—	—	—	—	—	—	—
Some high school	—	—	—	—	—	—	—	—
High school graduate	—	—	—	—	—	—	—	—
Some college	—	.2739	—	.4172	—	.2739	—	.4172
Age of immigration	.0724	.3463	.1042	.5214	.0724	.3463	.1042	.5214
Change in slope	.0037	.3500	.0002	.5216	.0039	.3502	.0004	.5218
Change in mean	.0003	.3503	.0001	.5217	.0002	.3504	.0005	.5223

Note. Analyses are based on data from the U.S. Bureau of the Census (1992). Samples consist of Spanish- and Chinese-speaking adults (age 18 or greater) who entered the United States between 1960 and 1980. Sample sizes: Spanish, n = 2,353,308; Chinese, n = 318,967.

entry variables in the language-group data we used (U.S. Bureau of the Census, 1995). Ideally, both of these variables would be reported at unit intervals, rather than by category. Variables in the language-group data were not reported at unit intervals; therefore, we completed an additional analysis to achieve greater specificity in testing relationships of interest. The Public Use Microdata Sample (PUMS; U.S. Bureau of the Census, 1992) provides exact age in years rather than categorical cross-tabulations, cutting significantly the possibility of "presmoothing."[2] Replicating the original analyses with the PUMS data (i.e., replacing the categorical age variable with its unit-level counterpart) added little additional insight; the incremental R^2 values were below 1%, providing little additional evidence for a critical period (as shown in Table 2).

[2]The PUMS data consist of a 1% sample of the entire census population, weighted to provide full population estimates.

SUMMARY

The existence of a critical period for second-language acquisition is the subject of much research and policy debate. Our original study found little support for a critical period—a finding questioned by Stevens. We have addressed each concern raised by Stevens; in each case, we have found little additional evidence to support the critical-period hypothesis. Stevens's methodological concerns with census-data analyses are well taken; researchers would be wise to keep these issues in mind in future treatments. However, in the current case, additional analyses addressing these concerns do little to suggest additional evidence for a critical period for second-language acquisition.

REFERENCES

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science*, *14*, 31–38.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied linear statistical models*. Chicago: Irwin.

Stevens, G. (2004). Using census data to test the critical-period hypothesis for second-language acquisition. *Psychological Science*, *15*, 215–216.

U.S. Bureau of the Census. (1992). *1990 public use microdata sample. 5-percent and 1-percent sample prepared for web download by the Bureau of the Census*. Retrieved January 6, 2004, from http://ftp2.census.gov/census_1990/1990_PUMS_A/

U.S. Bureau of the Census. (1995). *1990 census of population and housing: Detailed cross-tabulations of selected language groups for states: 1990* [CD-ROM]. Washington, DC: Bureau of the Census, Population Division.

(Received 1/30/04; Accepted 2/13/04)