Degree of Bilingualism and Cognitive Ability in Mainland Puerto Rican Children

Kenji Hakuta

Yale University

HAKUTA, KENJI. *Degree of Bilingualism and Cognitive Ability in Mainland Puerto Rican Children*. CHILD DEVELOPMENT, 1987, 58, 1372–1388. The relation between degree of bilingualism and cognitive ability was assessed longitudinally in low-income background Puerto Rican elementary school children in the United States. All subjects were enrolled in a transitional bilingual education program. 83 subjects, beginning in grades K–1, were followed for 3 years; 111 subjects, beginning in grades 4–5, were followed for 2 years. Cross-sectional and longitudinal models using regression procedures were tested for the hypothesis that degree of bilingualism is positively related to cognitive ability. Positive and statistically reliable results were obtained between nonverbal intelligence measures and degree of bilingualism in the younger cohorts, but the effects attenuated over time and age level. Metalinguistic awareness in the native language did not show a relation with degree of bilingualism. The results are interpreted in light of the sociolinguistic characteristics of the bilingual condition of the community.

The literature on childhood bilingualism contains a relatively large number of studies on its relation with various cognitive abilities, including metalinguistic awareness (Ben-Zeev, 1977; Cummins, 1976; Ianco-Worrall, 1972), cognitive flexibility (Peal & Lambert, 1962), and divergent thinking and creativity (Landry, 1974; Torrance, Wu, Gowan, & Allioti, 1970). In general, when bilingualism is defined as having equal facility in both languages (balanced bilingualism), bilinguals have demonstrated superior cognitive abilities when compared with monolingual controls matched on SES variables (see reviews of this literature in Diaz, 1983). Theories of why bilingualism might affect cognitive ability are still at a primitive state. One possibility is that an early objectification of language results from the use of two languages, leading to the superior use of verbal mediation to guide cognitive activity (see Hakuta, Ferdman, & Diaz, in press, for a review). With a few exceptions (Diaz, 1985; Duncan & De Avila 1979; Hakuta & Diaz, 1985) most of the studies of bilingualism and cognitive ability have been conducted outside of the educational context of the United States.

The major reason for the paucity of studies of bilingualism and cognitive ability in the United States has to do with the fact that balanced bilingualism is rarely found in the American educational context, especially among linguistic minorities. Rather, the policy of transitional bilingual education is to move the limited English proficient child as quickly as possible into monolingual English instruction (Hakuta, 1986). The success of a program is often evaluated exclusively on the basis of how much English is learned and how rapidly children can be exited from the program (Willig, 1985).

Theoretical considerations aside, even in terms of sheer demographics, the case can be made for the importance of conducting investigations of bilingualism and cognitive ability in the context of American bilingual education. It has been estimated that there were approximately 2.4 million children between the ages of 5 and 14 with limited proficiency in English, and this number is projected to increase to 3.4 million by the year 2000 (Oxford, Pol, Lopez, Stupp, Peng, & Gendell, 1980). At present, there is still very little

I thank my collaborators and assistants, Rafael Diaz (with whom I started this project), Sylvia Galambos, Bernardo Ferdman, Alicia Fernandez, Dalila Isern, Helen Kang, John McGowan, Juan Perez, and Luz Minerva Ramos. I also thank Aida Comulada and her staff and teachers of the bilingual program in New Haven under whom the data were collected. Grant support from the National Science Foundation (DAR-8010860) and the National Institute of Education (NIE-G-81-0123) helped collect the data. Preparation of this manuscript was supported in part by Contract 400-85-1010 from the Office of Educational Research and Instruction, U.S. Department of Education, for the Center for Language Education and Research. Requests for reprints should be mailed to Kenji Hakuta, Merrill College, University of California, Santa Cruz, CA 95060.

[*Child Development*, 1987, 58, 1372–1388. 0009-3920/87/5805-0007$01.00]

understanding of the development of their bilingualism and its relation with other cognitive abilities (McLaughlin, 1984).

The lack of knowledge is frequently aggravated by the political cloak on the debate on the education of language minorities. For example, one hears opinions about the research on the harmful effects of bilingualism (Thompson, 1952). Much of the literature on the negative consequences of bilingualism is based on the debate earlier this century between hereditarian and environmentalist interpretations of the low IQ performance of immigrant children, and is currently considered primarily of interest from the viewpoint of the history of science (Hakuta, 1986). Nevertheless, the studies are invoked in arguing against the development of both the native language and English in the schools.

In this article, we will describe one attempt to understand the process of the development of bilingualism and cognitive abilities in the context of a typical bilingual education context in the United States. A 3-year longitudinal study was conducted with Puerto Rican children in the bilingual education program in New Haven, Connecticut, during which assessments of their abilities in both languages, as well as their cognitive skills in metalinguistic, nonverbal, and social cognitive tasks were made.

The study differs from most previous investigations of the effect of bilingualism in children in three major respects. First, it is concerned with nonbalanced bilingual children who are in the process of becoming bilingual, rather than with balanced bilingual children. Second, it is conducted in the context of a social milieu that is best described as subtractive bilingualism (Lambert, 1975), in which the second language is developing in the community at the expense of the native language. And third, it is longitudinal in nature, allowing us to make some inferences about causality. The data from the first year of the study on some of the cognitive variables with the younger grade levels have been reported in several papers (Diaz, 1985; Hakuta & Diaz, 1985; Hakuta et al., in press). In this article, we report on the 3-year longitudinal aspect of the data and interpret the results in the broader context of bilingual education in the United States.

As in our previous studies, our operational definition of degree of bilingualism hinges on the particulars of the second language-learning situation of our subject population. As will be described below, the bilingual education program in the elementary school grades in New Haven maintains instruction in the native language to a considerable degree while the children acquire English. We can think of the bilingualism of our subject population as fitting somewhere in the two-dimensional space created by relative abilities in L1 and L2, as pictured in Figure 1*a*. This space is marked by a line with a diagonal line along which ideally balanced bilinguals will cluster. It should be noted that most previous studies of bilingualism in children that have looked at balanced bilingual children attempt to find subject populations that fall along this line. The two-space in Figure 1*a* is, of course, a theoretical idealization that is unmeasured, and we presently have no way of locating any given sample of bilingual individuals in absolute terms. Thus, depending on sample characteristics, the two-space defined by a set of L1 and L2 measures can vary. Figure 1*b* shows two hypothetical samples, A and B, where B represents a group of individuals that is more "balanced" than A. The point is that although we cannot determine the two-space location of a group of bilinguals in any particular study using a particular measure, the effects of degree of bilingualism can be studied by looking at the variation in the L2 measure while controlling for variation in the L1 measure. As can be seen in Figure 1*c*, individual A2 is more bilingual than individual A1, and B2 is more bilingual than B1. We hypothesize that the variation in L2 controlling for variation in L1 is attributable to the degree of bilingualism, and this variation should be related in a positive way to cognitive ability.

Method

Subjects

Subjects were from students in the bilingual education program in the New Haven public schools. Spanish is the only language for which bilingual services are provided in New Haven. Assignment to the bilingual program is determined on the basis of a combination of responses to a home language survey and teacher assessment of the student's skills in English. The policy of the program is to assign to the program only students who are dominant in Spanish and who are expected to be handicapped if they receive instruction exclusively in English. Standardized testing of English language proficiency is used for entry into the program when there is ambiguity about the language dominance of the student.

The program uses a "pairing model" instructional system. In this model, teachers

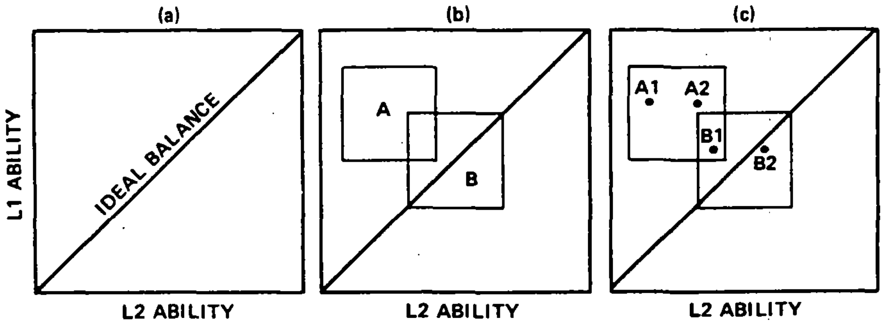

FIG. 1.—Conceptualization of balanced bilingualism, how different bilingual samples are embedded within it, and how individuals are embedded within samples. Panel *a* shows ideal balance line between L1 and L2 abilities. Panel *b* shows placement of two different samples (A and B) within idealized space. Panel *c* shows individuals within the sample spaces.

work in pairs, where one is primarily responsible for English and the other Spanish. Although many of the teachers are certified to teach in both languages, it is usually the case that the language assigned to the teacher is their native language. The pair of teachers is assigned two groups of students, who alternate between them. The most common model is one in which one group of children receives instruction in Spanish in the morning and instruction in English in the afternoon, while the second group receives English in the morning and Spanish in the afternoon. This characteristic of the program is important in that although the emphasis of the program is in the acquisition of English language skills, a substantial proportion of the instruction is still given in Spanish.

The Hispanic population in New Haven is primarily from Puerto Rico. In the elementary grades, as of October 1983, there were 1,652 Hispanic pupils in the New Haven public schools, comprising 20.1% of the entire elementary school body. Of these, roughly 48% (794) were in bilingual programs.

During the first year of data collection, we began observation at two grade levels: kindergarten (referred to as Cohort 0) and first grade (referred to as Cohort 1). This group, referred to collectively as Cohort 01, was followed over a 3-year period. Each year, we made two observations, once each in the fall and spring. Thus, Cohort 01 subjects who remained in the study for its duration were observed a total of six times.

We found considerable attrition over the years, due to the high mobility of our subject population. Out of a total of 155 subjects that we observed in the first testing period, only 83 remained throughout the 3 years. Although there are no firm data on the destination of the students who left the study, the teachers and staff of the program believe that a substantial proportion return to Puerto Rico. Although we added new subjects to the study as we went along in order keep up the numbers for cross-sectional analyses, this article will only report on the subjects for whom we have the complete set of longitudinal data.

During the second year, we also added cohorts of fourth and fifth graders (Cohort 4 and Cohort 5, respectively, Cohort 45 collectively), following the same schedule as Cohort 01, except that they were observed only over a 2-year period rather than 3. This cohort also saw considerable attrition. Out of a total of 152 subjects who began the study, 111 remained for the duration of the project.

In the longitudinal sample, the distribution of sex and the mean age at the beginning of the study was as follows: Cohort 0: $N = 38$ (18 girls, 20 boys), $M = 5.13$ years (SD = .41); Cohort 1: $N = 45$ (24 girls, 21 boys), $M = 6.60$ years (SD = .58); Cohort 4: $N = 65$ (37 girls, 28 boys), $M = 9.98$ years (SD = .78); Cohort 5: $N = 46$ (19 girls, 27 boys), $M = 10.76$ years (SD = .68).

Schools for sample selection were chosen by recommendation of the supervisor of the bilingual program. These were schools with highest concentrations of students in the bilingual classes. Within any given classroom, all students were initially screened through administration of a Spanish vocabulary test (described below). Students with low scores on the test (defined as greater than 1 SD below the group mean) were eliminated from our sample. This screening was considered important because of considerations related to special education programs for language minority students. There is often ambiguity as to the appropriate program assignment of language minority students with learning difficulties (Cummins, 1984). In the absence of bilingual special education programs, students of this category who happen to be of non-English home backgrounds may be assigned to bilingual education programs. In practice, our criteria were successful, in that they eliminated students on the low tail of a negatively skewed distribution. Teachers spontaneously commented on the success we had in identifying students with such difficulties. Five percent of the total group initially tested were eliminated from the sample in this manner.

The results of a bilingual questionnaire sent to the parents of our subjects, of which 77% were returned, revealed the following characteristics of their homes: An overwhelming majority of our subjects used only or mostly Spanish at home. For example, on a five-point scale ranging from 1 (only Spanish) to 5 (only English), the mean response to the question about what language was used by the children with adults at home was 2.03 (SD = .86) for Cohort 0, 2.00 for Cohort 1 (SD = .95), 2.11 (SD = .96) for Cohort 4, and 2.13 (SD = .99) for Cohort 5.

Median length of residence in the mainland United States was 96.5 months for Cohort 0, 96.3 months for Cohort 1, 108.8 months for Cohort 4, and 119.6 months for Cohort 5. Employment rate is extremely low in this group. The percentage of those who reported the head of household as being employed was 53.6% for Cohort 0, 35.1% for Cohort 1, 20.4% for Cohort 4, and 25.6% for Cohort 5. The mean number of adults in the household (M = 1.6 for Cohort 01, SD = .8, M = 1.5 for Cohort 45, SD = .8) indicates that a substantial percentage of the households have single parents. All subjects were eligible for the school lunch program, indicating the overall low socioeconomic condition of the students in the program. These quantitative data from the questionnaires are consistent with our observations and informal contacts with the parents, community, and schools in New Haven over the past 7 years.

Measures of Bilingualism

Estimates of relative abilities in L1 (Spanish) and L2 (English) were obtained through vocabulary tests. The English Peabody Picture Vocabulary Test (Dunn, 1965) and a Spanish translation adapted for Puerto Rican students in New York City (Wiener, Simmond, & Weiss, 1978) were chosen as the principal measures. This decision was based on several considerations.

First, after reviewing a large number of measures of language proficiency developed for use with Spanish-English bilingual elementary school children, none were judged appropriate for the range of age levels under study. Since the Peabody Picture Vocabulary Test was constructed for use with individuals from age 2-6 through 18-0, albeit monolingual English speakers, we felt that it contained the range of variation to be found in our subjects, for both English and for Spanish. Use of this test would be inappropriate in assigning mental age equivalents to our subjects. Rather, we were interested in the test's ability to assign relative abilities in both languages to our subjects.

Second, we had to deal with the practical problem of finding a test that could be administered in a short period of time, since we were testing our subjects individually and any test that took substantial time for administration would reduce the number of subjects that could be included in our study. Since the PPVT could be administered in approximately 20 min, we felt that it met these specifications.

Our use of the Spanish and English versions of the PPVT (hereafter SPVT and EPVT, respectively) as measures of proficiency in the languages was validated against independent measures of English and Spanish on a subset of our subjects. Scores on the vocabulary test compared well (r = .55 for Spanish, r = .62 for English, N = 49) with performance on the Language Assessment Scales (De Avila & Duncan, 1981), a proficiency measure for both English and Spanish commonly used with Spanish-speaking minority students by school systems. Vocabulary performance also correlated well with a measure of story-retelling fluency as rated by native-speaking judges (r = .82 for English, r = .36 for Spanish, N = 40).

Measures of Metalinguistic Awareness

Metalinguistic awareness refers to the ability to objectively analyze linguistic output. Different measures of metalinguistic awareness in Spanish were constructed for our younger and older cohorts. The full set of stimuli for these measures, as well as all of the measures used in this study, can be found in Hakuta (1984).

Cohort 01: metalinguistic awareness, Task A.—This task consisted of seven ungrammatical Spanish sentences with three correct sentences intermixed within the set. The sentences were read aloud, one at a time, and children were asked to decide whether the sentences were correctly said in Spanish or not. For those sentences that the subject judged to be ungrammatical, they were asked to provide a correct version of the sentence in Spanish. Children's corrections of the ungrammatical sentences were scored as (3) syntactic correction, (2) a combination of syntactic and semantic corrections, (1) semantic correction, and (0) no correction offered, on the assumption that correcting ungrammatical sentences on a syntactic rather than a semantic dimension indicates a higher awareness of the structural properties of language. A reliability of alpha = .83 was obtained for this measure.

Cohort 01: metalinguistic awareness, Task B.—The second metalinguistic measure consisted of presenting children seven Spanish sentences that contained one English word in them. Three correct Spanish sentences were randomly inserted within the set. Sentences were read aloud and, for each sentence, children were asked to judge them as correctly said in Spanish or not. Children's responses to each sentence were scored as (1) detecting bilingual mixing or (0) failure to detect mixed sentence. A reliability of alpha = .86 was obtained for this measure.

Cohort 45: ambiguity detection.—For the older cohort, it was determined that simple detection of ungrammatical sentences would be within most subjects' control. Based on the literature with monolingual children, we decided that detection of ambiguity in sentences would be appropriate for this age range. As in the metalinguistic tasks, we constructed various item types, whose differences we could test, but for the present report, we only report the total score on the entire test.

Subjects heard sentences played over a tape-recorder. After each sentence, the subject was asked how many meanings the sentence contained. The subject was then asked to paraphrase the meaning of each sentence. Finally, he or she was shown pictures that depicted the two meanings and asked whether the sentence could represent the pictures. In this article, we will report only the analysis based on the number of meanings actually produced by the subjects, which is the most preferable of the measures in that it is uncontaminated by response bias.

Nonverbal Measures

Raven's Coloured Progressive Matrices Test.—Subjects in all cohorts were tested on the Raven's Coloured Progressive Matrices Test (Raven, Court, & Raven, 1976). This was considered to be our primary measure of nonverbal cognitive ability, as it had been for Peal and Lambert's (1962) original work.

Thurstone's Primary Mental Abilities.—The Spatial Relations subtests of the Thurstone's Primary Mental Abilities Test were used. Grade-appropriate tests were used for Cohort 01 and Cohort 45. Two subparts (named here Task A and B) were administered for both cohorts. For both groups of cohorts, Task A involved choosing from four alternative geometric figures a figure that would complement a target figure to make up a square. The task for Cohort 45 was made more difficult, among other things, by having the figures rotated at an angle. In Task B for Cohort 01, the subject was asked to draw missing lines on a geometric figure to match a model figure. For Cohort 45, Task B involved grouping three of four complex geometric figures that were related on some logical basis. These spatial tests were not administered to Cohort 01 during their first year of testing.

Cohort 01: Chandler's bystander cartoons.—The present measure consists of a modified version of Chandler's bystander cartoons, originally devised as an individual differences measure of children's egocentrism (Chandler, 1973). The cartoons measure children's capacity to take the perspective of another person or, more specifically, the ability to separate their knowledge about a story from the knowledge of a bystander introduced in the middle of a story sequence.

Each child in the study was told two stories in a random order with the aid of cartoons. One story, "Sandcastle," portrayed a child whose sandcastle was destroyed by a girl riding a tricycle over it. The child then goes home and impulsively destroys his baby brother's card castle. Children in the study were asked to retell the story to the experimenter "from the point of view of the baby brother" or as "the baby saw it." Children were then asked three specific questions: (1) What is the baby thinking now? (2) Does the baby know why his brother destroyed his card castle? and (3) What does the baby think about his brother breaking his castle? Children's account of the story as well as their answers to the specific questions were scored as (0) egocentric response and (1) nonegocentric response, where egocentric responses reflected children's inability to separate their own knowledge about the story from the bystander's point of view. Similar procedures were followed for the second story.

Procedures

Subjects were tested individually, with the exception of the nonverbal measures in Cohort 45 after Time 3. It was assumed that for the older children, it would be more efficient to administer these measures in small groups of five or six children without affecting the results. In all other cases, testing was conducted by taking each child individually to a quiet part of the building. All testing except for the EPVT was conducted in Spanish by testers who were native speakers of Spanish.

In order to minimize the possibility of an experimenter bias, a testing schedule was created such that research assistants and testing sessions would not be confounded with specific tests. Most important, we made sure that in all cases, the EPVT and SPVT were administered by different research assistants on different days, such that the status of each subject in terms of his or her degree of bilingualism was kept blind to the testers. The remaining measures were administered approximating a counterbalanced order, although uneven numbers of subjects and practical considerations prevented a true counterbalanced design. However, we are confident that results would not be confounded with test order effects.

The EPVT, SPVT, Raven's, and the Spatial tests were scored immediately after test administration. For the metalinguistic tasks and for Chandler's, the sessions were tape-recorded and subsequently scored.

Results

Discussion of the results will be reported in two parts. In the first part, results are reported from analyses in which variables peripheral to the central scope of this article are analyzed. These include an analysis of the attrition of subjects from the study, an analysis of sex differences, and an analysis of practice effects from repeated testing. In the second part, we look at the relation between bilingualism and the cognitive measures. Simple cross-sectional models as well as longitudinal analyses are evaluated.

Since there were six observation periods for Cohort 01 and four for Cohort 45, for purposes of simplifying the longitudinal analysis as well as increasing the reliability of the measures, the data from the fall and spring testing of each year were averaged to create a yearly score. The single exception to this was for the analysis of the attrition effects, in which the data from the very first round of data collection were utilized to increase the size of the group that had dropped out of the study.

Peripheral Variables

Analysis of attrition effects.—In order to determine whether the subjects who remained for the duration of the longitudinal study differed from those who left the study, t tests were conducted between these two groups for the data collected in the fall of the first year. In Cohort 01, the longitudinal subjects showed a significantly higher mean in EPVT in the first testing session ($M = 27.17$) than those who left the study ($M = 21.01$), $t(153) = 2.68$, $p < .01$. The longitudinal group also showed a significantly higher mean in the Raven's, $t(149) = 2.15$, $p < .05$. There were no other significant differences on any of the other measures between the groups. In Cohort 45, the longitudinal group also showed a significant advantage over those who left the study in their EPVT, $t(150) = 2.41$, $p < .05$. There were no other significant differences. In general, then, the group in the longitudinal study were those who were more advanced in their English. This means that the effect of bilingualism for the longitudinal sample has a more restricted range than in the general population of students who are in bilingual education programs.

Analysis of sex differences.—In order to test for possible sex differences, analyses of variance were conducted for Cohort 01 and for Cohort 45, separately for each dependent measure in the study and treating sex and cohort as independent variables. For Cohort 01, sex was significant only in one instance of the metalinguistic awareness, Task A, $F(1,79) = 4.29$, $p < .05$, in favor of the girls. For Cohort 45, the analyses showed that boys were significantly higher than girls on the English measure at both times, $F(1,107) = 12.71$, $p < .001$, in the first year, and $F(1,107) = 8.26$, $p < .01$, in the second year. There were no significant sex × cohort interactions in any of the analyses. Although these sex effects may be the source of interest and speculation, particularly with respect to the differences in the acquisition of English in older cohorts, the paucity of sex differences in light of the numerous comparisons justifies the removal of sex as a variable from further discussion.

Analysis of practice effects.—An obvious concern with repeated testing in a study such as this one is that subjects become "wise" to the tests and contaminate the results. In order to test this possibility, the performance of Cohort 0 on all of the measures when they were in grades 1 and 2 was compared with Cohort 1 in those same grades. Similarly, Cohort 4 in grade 5 was compared with Cohort 5 in grade 5. The logic of these comparisons was that the cohorts were equivalent in terms of their grade level but differed in the number of times they had taken the test previously. Thus, when Cohorts 0 and 1 are compared in grade 1, if there is indeed a practice effect, Cohort 0 should have the advantage because they had taken the same test in kindergarten, while Cohort 1 was taking it for the first time.

The means between Cohort 0 and 1 at grades 1 and 2 showed no differences on t tests for independent samples. Inspection of the means showed no consistent trends. Thus, it could be concluded that there were no practice effects associated with the measures for Cohort 01.

Differences did crop up, however, for Cohort 45. The means for Cohort 4 were significantly higher than for Cohort 5 on EPVT (Cohort 4 $M = 70.05$, Cohort 5 $M = 63.28$, $t[109] = 4.56$, $p < .001$), on the Spatial Task B involving figure grouping (Cohort 4 $M = 17.59$, Cohort 5 $M = 16.39$, $t[109] = 2.18$, $p < .05$), and on the Ambiguity Task (Cohort 4 $M = 17.80$, Cohort 5 $M = 15.46$, $t[109] = 3.89$, $p < .001$). While these differences might be attributed to a practice effect, they could also be due to real differences in the two cohorts. At the higher grades, there is a rapid rate of mainstreaming students out of the bilingual program, such that students selected on the basis of being in the bilingual program at grade 4 and at grade 5 are different to the extent that selection out of the program between fourth and fifth grades is rapid. This possibility is underscored by the fact that while there was a difference on EPVT, there was no difference on SPVT, and mainstreaming is generally based on performance on English rather than Spanish.

Main Variables

The basic descriptive statistics for English, Spanish, and the dependent measures are presented separately for each cohort at each point in time in Table 1. The means show an expected increase with time for each cohort.

Cross-sectional models.—As we argued earlier, we conceptualized degree of bilingualism in this population as being the extent of development of the second language, taking into account the abilities in the native language. This was operationalized statistically as the variance in L2 (English) after the shared variance with L1 (Spanish) is removed. To evaluate the relation of bilingualism with cognitive ability, stepwise regression equations were set up separately for each cohort, for each variable, and for each time period in the following manner:

Cognitive Variable = Spanish/English.

The additional variance accounted for by adding English after Spanish had been included in the first step (ΔR^2) was tested for statistical significance.

TABLE 1

MEANS FOR ALL MEASURES OVER TIME, SEPARATELY BY COHORT

	COHORT 0 (N = 38)		
	Year 1 (Grade K)	Year 2 (Grade 1)	Year 3 (Grade 2)
EPVT	21.88 (13.08)	35.68 (12.41)	50.16 (12.57)
SPVT	40.07 (9.46)	56.25 (12.47)	64.04 (10.35)
Raven's	14.80 (2.11)	16.37 (2.72)	21.46 (3.74)
Metalinguistic, Task A	2.32 (2.82)	6.86 (4.34)	10.55 (4.59)
Metalinguistic, Task B	4.13 (1.75)	5.47 (2.17)	6.68 (1.65)
Spatial, Task A	...	8.28 (1.60)	10.03 (1.21)
Spatial, Task B	...	13.45 (1.80)	15.30 (1.75)
Chandler's cartoons	3.55 (1.20)	3.57 (1.76)	4.28 (1.52)

TABLE 1 (*Continued*)

		Cohort 1 (*N* = 45)	
	Year 1 (Grade 1)	Year 2 (Grade 2)	Year 3 (Grade 3)
EPVT	39.17 (11.87)	51.34 (10.56)	59.30 (8.46)
SPVT	55.14 (8.58)	64.71 (10.36)	70.47 (8.75)
Raven's	16.56 (3.05)	20.81 (4.03)	24.53 (4.23)
Metalinguistic, Task A	8.37 (4.37)	11.78 (3.75)	13.56 (4.08)
Metalinguistic, Task B	6.06 (1.77)	7.04 (1.46)	7.22 (1.49)
Spatial, Task A	...	9.68 (1.37)	10.63 (1.23)
Spatial, Task B	...	15.39 (1.48)	16.38 (1.17)
Chandler's cartoons	3.32 (1.49)	3.94 (1.78)	4.33 (1.73)
	Cohort 4 (*N* = 65)		
	Year 1 (Grade 4)	Year 2 (Grade 5)	
EPVT	58.95 (8.34)	70.05 (7.53)	
SPVT	79.42 (12.29)	86.34 (11.91)	
Raven's	24.71 (3.98)	26.94 (4.87)	
Spatial, Task A	12.62 (3.17)	13.86 (3.53)	
Spatial, Task B	16.31 (2.76)	17.59 (3.00)	
Ambiguity task	15.52 (3.27)	17.80 (2.98)	
	Cohort 5 (*N* = 46)		
	Year 1 (Grade 5)	Year 2 (Grade 6)	
EPVT	63.23 (8.10)	72.77 (7.88)	
SPVT	85.39 (17.23)	92.02 (16.29)	
Raven's	27.11 (4.14)	29.20 (3.80)	
Spatial, Task A	14.18 (3.06)	15.89 (3.55)	
Spatial, Task B	16.39 (2.64)	18.68 (2.40)	
Ambiguity task	15.46 (3.28)	18.40 (3.31)	

NOTE.—Standard deviations are in parentheses.

As an alternative, a second equation was tested in which the contribution of the native language, Spanish, was evaluated after English had been taken into account, as follows:

Cognitive Variable = English/Spanish.

The results of both analyses can be found in Table 2. In addition to the R^2 uniquely attributable to Spanish and to English, the multiple R^2 where both Spanish and English are included is also reported.

The overall pattern of results is quite clear. Bilingualism accounts for a significant proportion of the variance for Cohort 0 on the measures of nonverbal cognitive ability. This is true in grades K and 1 in the Raven's, and in grades 1 and 2 on Task A of the Spatial Test, and in grade 1 on Task B. However, this pattern is not replicated in Cohort 1.

Bilingualism does not account for any variance in the measures of metalinguistic

TABLE 2

Cross-sectional Models of Relations between Bilingualism and Cognitive Ability Measures

	Cohort 0		
	Grade K	Grade 1	Grade 2
Raven's:			
Native language (Spanish)	.139**	.052	.007
Bilingualism (English)	.124**	.219**	.033
Multiple R^2	.295**	.310***	.085
Metalinguistic awareness, Task A:			
Native language (Spanish)	.268***	.168**	.185***
Bilingualism (English)	.002	.024	.025
Multiple R^2	.278**	.216**	.422***
Metalinguistic awareness, Task B:			
Native language (Spanish)	.074$^+$	.186**	.146**
Bilingualism (English)	.002	.040	.006
Multiple R^2	.074	.256**	.271**
Spatial relations, Task A:			
Native language (Spanish)	...	.093*	.000
Bilingualism (English)	...	.146**	.117*
Multiple R^2	...	.280**	.169*
Spatial relations, Task B:			
Native language (Spanish)	...	.041	.000
Bilingualism (English)	...	.109*	.023
Multiple R^2	...	.174*	.038
Chandler's bystander cartoons:			
Native language (Spanish)	.065	.000	.004
Bilingualism (English)	.005	.039	.079*
Multiple R^2	.067	.039	.151*
	Cohort 1		
	Grade 1	**Grade 2**	**Grade 3**
Raven's:			
Native language (Spanish)	.005	.035	.185***
Bilingualism (English)	.065*	.020	.013
Multiple R^2	.089	.127*	.352***
Metalinguistic awareness, Task A:			
Native language (Spanish)	.121*	.160**	.010
Bilingualism (English)	.002	.007	.029
Multiple R^2	.126*	.190**	.080
Metalinguistic awareness, Task B:			
Native language (Spanish)	.017	.050	.109*
Bilingualism (English)	.069*	.018	.033
Multiple R^2	.117*	.150*	.109*
Spatial relations, Task A:			
Native language (Spanish)	...	.015	.020
Bilingualism (English)	..	.002	.045
Multiple R^2	...	.034	.137*

TABLE 2 (*Continued*)

	Cohort 1		
	Grade 1	Grade 2	Grade 3
Spatial relations, Task B:			
Native language (Spanish)	...	.068*	.021
Bilingualism (English)	...	.010	.014
Multiple R^2	...	.071	.075
Chandler's bystander cartoons:			
Native language (Spanish)	.047	.018	.064*
Bilingualism (English)	.048	.024	.000
Multiple R^2	.073	.096	.089

	Cohort 4	
	Grade 4	Grade 5
Raven's:		
Native's language (Spanish)	.023	.009
Bilingualism (English)	.042*	.038
Multiple R^2	.115*	.081*
Spatial, Test A:		
Native language (Spanish)	.021	.078*
Bilingualism (English)	.008	.001
Multiple R^2	.050	.090*
Spatial, Test B:		
Native language (Spanish)	.009	.010
Bilingualism (English)	.054*	.021
Multiple R^2	.054	.056
Ambiguity detection:		
Native language (Spanish)	.216***	.129**
Bilingualism (English)	.005	.005
Multiple R^2	.237***	.200***

	Cohort 5	
	Grade 5	Grade 6
Raven's:		
Native language (Spanish)	.005	.000
Bilingualism (English)	.015	.005
Multiple R^2	.016	.006
Spatial, Test A:		
Native language (Spanish)	.030	.012
Bilingualism (English)	.008	.000
Multiple R^2	.031	.015
Spatial, Test B:		
Native language (Spanish)	.167**	.009
Bilingualism (English)	.043	.010
Multiple R^2	.176**	.032
Ambiguity detection:		
Native language (Spanish)	.049	.310***
Bilingualism (English)	.000	.001
Multiple R^2	.054	.405***

NOTE.—Numbers on rows marked "Native language" are R^2 changes in predicting cognitive ability with Spanish after English has been entered in the first step of the hierarchical regression. Numbers on rows marked "Bilingualism" are R^2 changes due to English after Spanish has been entered in the first step. All significant R^2s have positive *b* values.

$^+ p < .10$.

$^* p < .05$.

$^{**} p < .01$.

$^{***} p < .001$.

awareness in either Cohorts 0 or 1, with the exception of a marginally significant effect for Cohort 1 in grade 1. Rather, metalinguistic awareness is strongly related to native language ability.

The social-cognitive measure, Chandler's Bystander Cartoons, was unrelated to either bilingualism or to native language ability. This was perhaps due to the lack of sensitivity of our measure to social-cognitive skills. As can be seen from an inspection of the multiple R^2s, this measure was related in a significant way to the language measures only for Cohort 0 in grade 2.

The results from Cohort 45 were sporadic. The nonverbal measures were not related to bilingualism, with the possible exception of two marginally significant effects for Raven's and Spatial Test B in Cohort 4 at grade 4. The metalinguistic task of ambiguity detection was strongly related to native language ability but not to bilingualism. The pattern suggests, then, that bilingualism is related to nonverbal cognitive skills in the younger cohort, but this effect disappears in the higher grades.

Longitudinal models.—Due to limitations of small sample size in each of the cohorts, complex longitudinal analyses that would evaluate the fit of the data with specified factor structures, such as those employing LISREL, were avoided. Rather, hierarchical regression models taking advantage of the longitudinal nature of the data set were tested.

The logic of the longitudinal models was simple. It was assumed that if bilingualism at an earlier point in time could predict performance on the cognitive measures at a later point in time, this would imply support for the argument that bilingualism plays some causal role in the capacity underlying performance on the cognitive measure. On the other hand, if cognitive ability at an earlier point in time could predict bilingualism at a later point in time, it would suggest that superior cognitive ability leads to higher levels of bilingualism. Regression models were set up to evaluate each of these possibilities separately.

Two tests of this longitudinal logic were made. The first might be called a "weak" longitudinal model. In this model, the correlation of any given measure with itself over time was left out of the equation. Thus, to test the ability of bilingualism at an earlier point in time to predict cognitive ability at a later point in time, the following equation was employed:

Cognitive Variable (later) = Spanish (earlier)/
English (earlier).

This weak model takes only the cognitive ability at a later point in time. The predictive ability of earlier bilingualism on later cognitive ability does not take into account cognitive ability at the earlier point in time. Thus, it is a weak model in that it fails to account for *change* in cognitive ability between the earlier and later points in time, which presumably could be due to the force exerted by earlier bilingualism. Thus, a stronger model was set up, in which the covariance of the later cognitive ability with earlier cognitive ability was also removed in the first step of the equation, as follows:

Cognitive Variable (later) = Cognitive Variable (earlier)
+ Spanish (earlier)/English (earlier).

The alternative equations in which later bilingualism is predicted by earlier cognitive ability were set up by switching the two variables. Thus, the weak longitudinal model was as follows:

English (later) = Spanish (later)/
Cognitive Variable (earlier).

The stronger longitudinal model was as follows:

English (later) = English (earlier)
+ Spanish (later)/Cognitive Variable (earlier).

In these analyses, it should be recalled that there were three separate time periods of observation for Cohort 01, and two for Cohort 45. In both cases, time periods were analyzed in pairs. Thus, for Cohort 01, there were three separate pairings of the time periods. Year 1 was evaluated with Year 2, then Year 2 with Year 3, then Year 1 with Year 3. For Cohort 45, Year 1 was evaluated with respect to Year 2. In all cases, the weak model was tested first. The strong longitudinal model was tested only for instances where the weaker causal model yielded a significant result.

The results appear in Table 3. As in the cross-sectional analyses, most of the significant results appear in Cohort 0, concentrated around the nonverbal measures of Raven's and the Spatial Tasks. For Raven's, the weak longitudinal model between grades K and 1 works in both directions. Thus, bilingualism at kindergarten predicts a significant amount of variance in Raven's at first grade ($\Delta R^2 = .147, p < .001$), and Raven's at kindergarten predicts a significant amount of variance in bilingualism at first grade ($\Delta R^2 = .254, p < .001$). Tests of the strong models of this relation still maintained some robustness as can be seen in Table 3. For Raven's between grades 1 and 2, the model in which Raven's at first grade predicts bilingualism at second grade showed some strength.

For the spatial tasks, the direction of the effect also seemed to go in both directions. Thus, for example, bilingualism in kindergarten predicted performance on Task A in first grade, while performance on Task A predicted degree of bilingualism in second grade. A similar pattern can also be observed in Task B.

It is of some interest that the measures that did not show much relation with bilingualism in the cross-sectional analyses revealed some relation in the longitudinal analyses. Bilingualism at grade 1 predicted performance on Task A of the metalinguistic awareness measure at grade 2, and this relation was robust both in the weak ($\Delta R^2 = .152$, $p < .01$) and the strong ($\Delta R^2 = .087$, $p < .01$) models. The social-cognitive measure, Chandler's *cartoons*, also showed some relation with earlier measures of bilingualism, as can be seen in Table 3.

In Cohort 1, as in the cross-sectional analyses, there was little of statistical significance to report. However, where there were significant relations, the direction seemed to be in the direction of earlier bilingualism predicting later cognitive ability. In the weak model, bilingualism at first grade significantly predicted Raven's at third grade ($\Delta R^2 = .102$, $p < .01$), although this effect was attenuated when the strong model was tested ($\Delta R^2 = .037$, $p < .10$). A similar trend can also be seen by inspecting the results for Spatial Task A.

Cohort 4 showed some effects of earlier bilingualism predicting later cognitive ability. The weak model fared well for Raven's ($\Delta R^2 = .13$, $p < .01$), and was robust even in the strong model ($\Delta R^2 = .053$, $p < .01$). In the Spatial Task B, there was also a similar trend ($\Delta R^2 = .085$, $p < .01$, for the weak model, $\Delta R^2 = .024$, $p < .10$, for the strong model). Cohort 5 showed no significant relations on any of the models tested.

Relation of English and Spanish

In order to help interpret these results, the correlations between the English and the Spanish measures over time are shown in Table 4. The relation between the two languages increases over time in Cohort 01. For example, in Cohort 0, the correlation between the languages goes from $r = .11$ (N.S.) in kindergarten, to $r = .14$ (N.S.) in grade 1, to $r = .56$ ($p < .001$) in grade 2. In Cohort 1, the correlation starts at $r = .31$ ($p < .05$) in grade 1, $r = .57$ ($p < .001$) in grade 2, and $r = .54$ ($p < .001$) in grade 3. As can be seen in Table 3, in Cohort 45, the correlation is in the $r = .40$ range. Thus, in the early grades, the two languages are uncorrelated, but the relative abilities in the two languages become similar over time.

Discussion

The results indicated some positive correlations between bilingualism and the nonverbal measures of cognitive ability. The correlations were more consistent in kindergarten and first grade, with attenuation in the higher grades. The attenuation of the effect could either be due to changes in the reliability of the measures in the higher grades, or for the theoretically richer reason that the most interesting developments in the relation between cognitive ability and bilingualism are occurring in the early grades.

The longitudinal analyses suggest that some interesting interactions are occurring at the early stages. For the Raven's, there was some support for the effect going in both directions, that is to say, earlier performance in Raven's predicted later degree of bilingualism, as well as an earlier degree of bilingualism predicting later performance on Raven's. The longitudinal models for the spatial tasks could not be fully tested due to the fact that these dependent measures were not administered during the first year of the study. However, the results were consistent with those obtained for the Raven's. These results suggest that perhaps the alternative models about the direction of causality are not mutually exclusive, and that the relation between bilingualism and cognitive ability is an interactive one.

The metalinguistic awareness measures showed a consistently strong and positive relation with Spanish, but there was little evidence showing a relation with bilingualism. In the longitudinal analyses, there was one statistically reliable result between first and second grade for Cohort 0, going from bilingualism to metalinguistic performance. It can be concluded that our measures of metalinguistic ability failed to show any consistent relations with bilingualism.

Perhaps useful in interpreting this result is Bialystok and Ryan's (1985) distinction between linguistic knowledge and cognitive control in conceptualizing metalinguistic awareness, a distinction deriving from Shiffrin and Schneider (1977). The measure that

TABLE 3

CHANGES IN R^2 VALUES AND STATISTICAL SIGNIFICANCE OF LONGITUDINAL MODELS

	Cohort 0		
	K → 1	1 → 2	K → 2
Raven's:			
BILING → Raven's	.147***/.051+	.053	.014
Raven's → BILING	.254***/.037*	.129**/.018	.043
Metalinguistic awareness, Task A:			
BILING → MetaA	.015	.152**/.087**	.037
MetaA → BILING	.014	.026	.010
Metalinguistic awareness, Task B:			
BILING → MetaB	.004	.063*	.028
MetaB → BILING	.008	.003	.000
Spatial relations, Task A:			
BILING → SpatA	.164**/a	.059	.042
SpatA → BILING	b	.115**/.026	b
Spatial relations, Task B:			
BILING → SpatB	.090+	.031	.052
SpatB → BILING	b	.067*	b
Chandler's bystander cartoons:			
BILING → Chandler	.122*/.124*	.108*/.047+	.092+/.091+
Chandler → BILING	.077*	.052*	.015
	Cohort 1		
	1 → 2	2 → 3	1 → 3
Raven's:			
BILING → Raven's	.040	.020	.102**/.037+
Raven's → BILING	.007	.021	.045
Metalinguistic awareness, Task A:			
BILING → MetaA	.002	.038	.001
MetaA → BILING	.000	.032	.004
Metalinguistic awareness, Task B:			
BILING → MetaB	.046	.015	.013
MetaB → BILING	.053+	.006	.038
Spatial relations, Task A:			
BILING → SpatA	.009	.061+	.097*/a
SpatA → BILING	b	.001	b
Spatial relations, Task B:			
BILING → SpatB	.024	.027	.022
SpatB → BILING	b	.002	b
Chandler's bystander cartoons:			
BILING → Chandler	.015	.015	.037
Chandler → BILING	.000	.000	.000
	Cohort 4		
	4 → 5		
Raven's:			
BILING → Raven's	.130**/.053**		
Raven's → BILING	.008		
Spatial, Test A:			
BILING → SpatA	.010		
SpatA → BILING	.022		
Spatial, Test B:			
BILING → SpatB	.085**/.024+		
SpatB → BILING	.008		
Ambiguity detection:			
BILING → Ambig	.020		
Ambig → BILING	.003		

TABLE 3 (*Continued*)

	Cohort 5
	5 → 6
Raven's:	
BILING → Raven's	.000
Raven's → BILING	.001
Spatial, Test A:	
BILING → SpatA	.003
SpatA → BILING	.014
Spatial, Test B:	
BILING → SpatB	.000
SpatB → BILING	.021
Ambiguity detection:	
BILING → Ambig	.029
Ambig → BILING	.003

NOTE.—The weak model was tested in all instances where data were available. The strong model was tested only if the weak model was statistically reliable. When the strong model was tested, the change in R^2 is reported following the slash (/). All significant R^2s have positive b values.

[a] The strong longitudinal model could not be tested because the cognitive measure was not administered during Year 1 of the study.

[b] The weak longitudinal model could not be tested because the cognitive measure was not administered during Year 1 of the study.

$^+ p < .10$.

$^* p < .05$.

$^{**} p < .01$.

$^{***} p < .001$.

we used was very sensitive to the subjects' proficiency in Spanish, and it may well indicate that they were more sensitive to *knowledge* about Spanish, rather than the ability the *control* over cognitive activities on this knowledge. Indeed, such an interpretation is favored in our analysis from a psycholinguistic perspective of the specific item types employed in the measures (Galambos & Hakuta, 1986).

On the other hand, if it turns out to be the case that metalinguistic ability was being well tapped by our tasks, the result is rather disturbing from a theoretical point of view. Why should becoming bilingual, a primarily linguistic activity, have an impact on nonverbal cognitive activities and not on metalinguistic performance? We would either be forced to reconceptualize what it means to become bilingual, or minimally to draw less of a clear-cut distinction between verbal and nonverbal processes.

The relation between bilingualism and our measure of social perspective taking turned up some suggestive data, particularly in the longitudinal analyses. The model testing for earlier bilingualism predicting later performance on Chandler's was supported. There were also some indications of the effects in the other direction as well. This measure was perhaps the weakest of all that we used in the study, since the score range was 6. Given the difficulty of measuring social skills to begin with, the lack of very strong effects is not surprising. However, the results suggest a role that social-cognitive skills may play in the development of bilingualism, particularly in the area of code switching (Gumperz, 1982).

The overall results of this study must be interpreted within the limits defined by the subject population being investigated. A major limit was a high rate of attrition during the 3 years of the study due to the mobility of the Puerto Rican population in New Haven. In the younger cohort, subjects who left the study were significantly lower in their English proficiency as well as their Ravens performance than those who remained. Thus, the longitudinal sample represented the higher end of the distribution of these characteristics. This selective retention in the study resulted in a more restricted range in these variables. In the higher cohort, there was also a

TABLE 4

CORRELATION BETWEEN EPVT AND SPVT ACROSS TIME, SEPARATELY BY COHORT

SPVT	EPVT; COHORT 0 (N = 38)		
	Grade K	Grade 1	Grade 2
Grade K	.107	.135	.328*
Grade 1	.047	.147	.420**
Grade 2	.188	.355*	.564***
	EPVT; COHORT 1 (N = 45)		
	Grade 1	Grade 2	Grade 3
Grade 1	.307*	.362**	.364**
Grade 2	.595***	.566***	.530***
Grade 3	.523***	.505***	.540***
	EPVT; COHORT 4 (N = 65)		
	Grade 4	Grade 5	
Grade 4	.441***	.521***	
Grade 5	.328**	.455***	
	EPVT; COHORT 5 (N = 46)		
	Grade 5	Grade 6	
Grade 5	.289*	.429**	
Grade 6	.319*	.433**	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

significant difference in English in the same direction, though not on any of the other measures.

Another limit was that even at the end of the study, the subjects had not achieved a level of proficiency in English equivalent to what would be expected of balanced bilinguals. Even though the English PPVT norms should be used only with great caution because they were standardized with a different population, the mean raw scores corresponded to low age equivalents. The relatively weak effects observed in this study may be attributable to the fact that the subjects had not yet attained a state of balanced bilingualism (Cummins, 1976). Considering the sociolinguistic circumstances of the community, the majority of our children probably never will attain this state.

The most important constraint in helping to understand and generalize the findings of this study has to do with the population characteristics of the bilingual group studied, in combination with the schooling policy for bilingual students. The bilingualism of the Puerto Rican community in New Haven as a whole can be described as subtractive, where Spanish is being replaced by English (Hakuta, Ferdman, & Diaz, in press). Yet contrary to the trend in the community at large, the students, while they are in the bilingual program, are in an additive setting. Although many bilingual education programs in the United States have very limited support in the native language of the students, with a very heavy emphasis on rapid transition to monolingual English, the New Haven program places a strong emphasis on the development of the basic skills in the native language. Thus, the second language, English, is added as something of an enrichment to the students.

However, the goal of the New Haven program, like most bilingual education programs in the United States, is transition to monolingual English rather than maintenance of the native language. Thus, this enrichment in the additive context lasts only while the students are in the program, which for the average student corresponds to about 3 years. When the students are exited from the program, they follow the general subtractive forces of the community. It is an empirical question for future research to determine the extent to which these results are influenced by the conditions of additive and subtractive bilingualism.

Also of interest for further inquiry is the nature of the longitudinal relation between the native language and the second language. The increasing correlation between the two languages over time is suggestive, and supportive of the notion that the native language provides a foundation for second language acquisition. At the same time, it is worth considering the implication of the increasing correspondence between Spanish and English if the children were becoming bilingual in a truly additive sociolinguistic setting. If we were to make projections in this hypothetical world, the two languages would become increasingly correlated, to the point where, in the limit, they would become balanced bilinguals. To the extent that English is correlated with nonverbal types of measures, and Spanish is correlated to metalinguistic awareness measures, it is likely that balanced bilinguals, particularly where balanced bilinguals are those selected on the basis of strong verbal ability, will be those who are good on both kinds of cognitive abilities.

Finally, despite all of the discussion about the effects of bilingualism, it is worth considering what our measure of bilingualism really reflects, for essentially it is the degree of acquisition of English (controlling for basic ability in Spanish). If one thinks of the fact that the children are learning English primarily in the classrooms, then the extent to which English is learned might be thought of as a good measure of how well they are able to learn in general in the classroom setting. To the extent that our nonverbal measures were tapping this general ability to learn as well, it would not be surprising to find correlations with English. This possibility of a "third variable" of general brightness of the students producing the results cannot be ruled out in the absence of a true experiment. Given that bilingualism is not distributed in a randomly assignable fashion, we will simply have to learn to live with and understand this confound.

This study, while testing some specific hypotheses about the role of bilingualism in the cognitive ability of children, ran into practical difficulties that are part of the realities of research with many language minority populations. These include high mobility and the selectivity of students who are placed in the program. The cohort differences that were observed in this study are undoubtedly due in part to perturbations introduced by these characteristics of the population. As we investigated the relatively cognitive and linguistic angles of development in this population, it became increasingly clear that this could not be isolated from the patterns of language maintenance and use in this population as they relate to its mobility and other demographic patterns.

References

Ben-Zeev, S. (1977). The influence of bilingualism on cognitive strategy and cognitive development. *Child Development*, **48**, 1009–1018.

Bialystok, E., & Ryan, E. B. (1985). Toward a definition of metalinguistic skill. *Merrill-Palmer Quarterly*, 31, 229–251.

Chandler, M. J. (1973). Egocentrism and antisocial behavior: The assessment and training of social perspective taking skills. *Developmental Psychology*, 9, 326–332.

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism*, 9, 1–43.

Cummins, J. (1984). *Bilingualism and special education*. San Diego: College Hill.

De Avila, E. A., & Duncan, S. E. (1981). *Language assessment scales*. San Rafael, CA: Linguametrics.

Diaz, R. M. (1985). Thought and two languages: The impact of bilingualism on cognitive development. *Review of Research in Education*, **10**, 23–54.

Duncan, S. E., & De Avila, E. A. (1979). Bilingualism and cognition: Some recent findings. *NABE Journal*, 4, 15–50.

Dunn, L. (1965). *Peabody Picture Vocabulary Test*. Circle Pines, MN: American Guidance Service.

Galambos, S., & Hakuta, K. (1986). *Subject-specific and task-specific characteristics of metalinguistic awareness in bilingual children*. Unpublished manuscript, Department of Psychology, Yale University.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.

Hakuta, K. (1984). *The causal relationship between the development of bilingualism, cognitive flexibility, and social-cognitive skills in Hispanic elementary school children*. Final Report, National Institute of Education Grant G-81-0123. Available from the National Clearinghouse for Bilingual Education, Arlington, VA.

Hakuta, K. (1986). *Mirror of language: The debate on bilingualism*. New York: Basic.

Hakuta, K., & Diaz, R. M. (1985). The relationship between degree of bilingualism and cognitive ability: A critical discussion and some new longitudinal data. In K. E. Nelson (Ed.), *Children's language* (Vol. 5, pp. 319–344). Hillsdale, NJ: Erlbaum.

Hakuta, K., Ferdman, B. M., & Diaz, R. M. (in press). Bilingualism and cognitive development: Three perspectives. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics: Vol. 2. Reading, writing and language learning*. New York: Cambridge University Press.

Ianco-Worrall, A. (1972). Bilingualism and cognitive development. *Child Development*, **43**, 1390–1400.

Lambert, W. E. (1975). Culture and language as factors in learning and education. In A. Wolfgang (Ed.), *Education of immigrant students* (pp. 55–83). Toronto: Ontario Institute for Studies in Education.

Landry, R. G. (1974). A comparison of second language learners and monolinguals on divergent thinking tasks at the elementary school level. *Modern Language Journal*, **58**, 10–15.

McLaughlin, B. (1984). *Second language acquisition in childhood: Vol. 1. Preschool children*. Hillsdale, NJ: Erlbaum.

Oxford, R., Pol, L., Lopez, D., Stupp, P., Peng, S., & Gendell, M. (1980). *Projections of non-English language background and limited English proficient persons in the United States to the year 2000*. Rosslyn, VA: InterAmerica Research Associates.

Peal, E., & Lambert, W. E. (1962). The relation of

bilingualism to intelligence. *Psychological Monographs*, 76(546), 1–23.
Raven, J. C., Court, J. H., & Raven, J. (1976). *Raven's Progressive Matrices Test*. London: H. K. Lewis.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: 2. Perception, learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–190.
Thompson, G. G. (1952). *Child psychology*. Boston: Houghton Mifflin.
Torrance, E. P., Wu, J. J., Gowan, J. C., & Allioti, N. C. (1970). Creative functioning of monolingual and bilingual children in Singapore. *Journal of Educational Psychology*, 61, 72–75.
Wiener, F. D., Simmond, A. J., & Weiss, F. L. (1978). *Spanish Picture Vocabulary Test*. New York: Marymount Manhattan College.
Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, 55, 269–317.