Subject-specific and task-specific characteristics of metalinguistic awareness in bilingual children

SYLVIA JOSEPH GALAMBOS and KENJI HAKUTA
Yale University

ADDRESS FOR CORRESPONDENCE
Board of Studies in Education, Merrill College, UCSC, Santa Cruz, CA 95064

ABSTRACT
The relationship between bilingualism and metalinguistic awareness was explored in Puerto Rican Spanish- and English-speaking children. All subjects were from low-income backgrounds and were enrolled in a transitional bilingual education program in the United States. Two longitudinal studies were conducted. The first study examined the abilities to note and correct ungrammatical sentences in Spanish. Subjects were 104 children in first and second grade at the beginning of the study. They were followed over a period of 2 years. The second study looked at the ability to detect ambiguity in sentences, and to paraphrase the different meanings. There were 107 subjects who were in fourth and fifth grades, and were also followed over a 2-year period. The results from both studies indicated that native language proficiency as well as the degree of bilingualism affected metalinguistic awareness. The results also indicated that these effects interacted with the types of items in the metalinguistic tasks. This suggests that both subject-specific and task-specific factors are important in understanding the relationship between bilingualism and metalinguistic awareness.

The goal of this paper is to explore the relationship between bilingualism and metalinguistic awareness – the ability to attend and reflect upon the properties of language. This relationship has been studied by a number of previous researchers who, for the most part, have hypothesized that bilingualism accelerates the conscious separation of the name from the object to which it refers (Ben-Zeev, 1977a, 1977b; Cummins, 1978; Ianco-Worrall, 1972; Rosenbaum & Pinker, 1983). Although this hypothesis has been supported in some experiments, it has not been in others (see, e.g., Ben-Zeev, 1977; Cummins 1976; see Galambos & Goldin-Meadow, 1986, for a review). The discrepancy in these results might be due to differences in the populations studied, such as in their age, level of proficiency in the two languages, and the context of bilingualism. These differences, however, are difficult to assess, given the lack of detailed information about the populations.

The concern of our study was two-fold. We wanted to test a hypothesis we believe to be more promising – that the bilingual experience, by requiring that the form of the two languages be attended to, enhances the ability to direct attention to the form of language and focus on it effectively. Our hypothesis,

therefore, is that bilingualism does not necessarily enhance knowledge about the properties of language, but rather alters the *approach* toward language, from a message-oriented approach to a form-oriented one.

Our second concern was to explore the relationship of subject-specific factors, such as degree of bilingualism, and task-specific factors, such as the particular types of items tested, to metalinguistic awareness. *Bilingualism* has been used as a cover term for very differing populations, and *metalinguistic awareness* has been examined globally rather than psycholinguistically. We need to develop a more detailed picture of both notions before being able to make conclusive claims concerning their relationship.

In this paper we will report the results of two studies conducted with low-income Spanish- and English-speaking bilingual children in the Bilingual Education Program in the New Haven Public Schools. Assignment to the bilingual program is determined on the basis of a combination of responses to a home language survey and the teacher's assessment of the student's skills in English. The policy of the program is to assign to the program only students who are highly dominant in Spanish and who are unable to benefit from instruction in English. Standardized testing of English-language proficiency is used for entry into the program only when there is ambiguity about the language dominance of the student. Although the emphasis of the program is in the acquisition of English-language skills, a substantial proportion (roughly half) of the instruction is still given in Spanish. Children are mainstreamed into English-only classrooms when it is judged that they are proficient enough in English to benefit from English-only instruction. Two out of three criteria must be met before main-streaming. Performance on the Language Assessment Battery (developed by the State of New York) must meet the exiting criterion that corresponds to a score at or above the 30th percentile, with reference to Hispanic students primarily in New York State. Teachers' assessments of proficiency in English and academic performance are the other criteria.

The children in the first study were first- to third-grade children and those in the second were fourth- to sixth-grade children. Because many of the population characteristics were similar for the two groups of children, the subject section below will contain information on both groups. The materials, procedure, and results will be described separately for the two studies. The metalinguistic tasks chosen for the two studies were felt to tap similar metalinguistic skills and, as indicated by previous research with monolingual populations, were age appropriate. We will discuss this further below.

Subjects

The subjects in Experiment 1 were 52 first-grade children (Cohort 1), tested 4 times over a 2-year period, twice in first grade and twice in second grade, and 52 second-grade children (Cohort 2), tested four times as well, twice in second grade and twice in third grade. Each year, we made two observations, once in the fall and once in the spring. If a child had been mainstreamed into English-only classrooms during the testing period, he was followed into these classrooms for testing. Thus, although the 104 children selected were all initially in the bilingual

Table 1. *Mean age and range for each cohort*

Cohort 1	$n = 52$	Mean age at T1 (1st grade)	= 6;5	Range:	5;0– 9;0.
		Mean age at T2 (1st grade)	= 6;11	Range:	5;6– 9;6.
		Mean age at T3 (2nd grade)	= 7;5	Range:	6;0–10;0.
		Mean age at T4 (2nd grade)	= 7;11	Range:	6;6–10;6.
Cohort 2	$n = 52$	Mean age at T1 (2nd grade)	= 7;6	Range:	6;1– 9;3.
		Mean age at T2 (2nd grade)	= 8;0	Range:	6;7– 9;6.
		Mean age at T3 (3rd grade)	= 8;6	Range:	7;1–10;0.
		Mean age at T4 (3rd grade)	= 9;0	Range:	7;7–10;6.
Cohort 4	$n = 63$	Mean age at T1 (4th grade)	= 10;3	Range:	9;2–12;2.
		Mean age at T2 (5th grade)	= 11;3	Range:	10;2–13;2.
Cohort 5	$n = 44$	Mean age at T1 (5th grade)	= 11;0	Range:	9;9–12;6.
		Mean age at T2 (6th grade)	= 12;0	Range:	10;9–13;6.

program, by the second year about 20% of the children had been mainstreamed into English-only classrooms.

The subjects in Experiment 2 were 63 fourth-grade children (Cohort 4), tested twice, the first time in fourth grade and the second time in fifth grade, with a year's separation between testing, and 44 fifth-grade children (Cohort 5), also tested twice, once in fifth grade and once in sixth grade. If a child had been mainstreamed, he was also followed into the English-only classrooms for testing. Approximately 45% of the children in both cohorts had been mainstreamed by the second testing time.

The initial sample size for Cohorts 1 and 2 combined was 150 children and for Cohorts 4 and 5, 185 children. However, considerable attrition due to the high mobility of the subject population reduced the sample size of both groups considerably. In addition, the initial subject population was screened for Spanish proficiency through the administration of a Spanish translation of the English Peabody Picture Vocabulary Test (see description of measure under discussion of materials and procedure for Experiment 1). Students with low scores on the test (defined as greater than one standard deviation below the group mean) were eliminated from the sample. Our motivation for this was to avoid including language minority children with learning difficulties, who in the absence of bilingual special education programs, may be assigned to bilingual education programs. Five percent of the total group initially tested were eliminated from the sample in this manner.

The mean age for Cohort 1 at the first testing time was 6;5 (years;months) and the age range was 5;0 to 9;0 years of age. The mean age for Cohort 2 at the first testing time was 7;6, with an age range of 6;1 to 9;3. The age information for the four testing times is displayed in Table 1 for the two cohorts. There were 25 females and 27 males in Cohort 1 and 27 females and 25 males in Cohort 2. Point-biserial correlations between sex and any of the dependent measures were insignificant ($r < .10$).

The mean age for the subjects in Experiment 2 was 10;3 for Cohort 4 at the first testing time, with an age range of 9;2 to 12;2; for Cohort 5, the mean age

was 11;0, with an age range of 9;9 to 12;6. Information on the age of the subjects in this experiment can also be found in Table 1. There were 35 females and 28 males in Cohort 4, and 19 females and 25 males in Cohort 5. Sex did not correlate significantly with any of the dependent measures in these cohorts.

The subjects in both experiments were mostly of Puerto Rican origin; 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. A questionnaire sent to the parents of our subjects, of which 76% 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 5-point scale ranging from 1 = only Spanish to 5 = only English, the mean response on the language used by the children with adults at home was 1.9 ($s = .88$) for Cohorts 1 and 2 combined and 2.06 ($s = .94$) for Cohorts 4 and 5. Mean length of residence in the mainland United States was 90.1 months for Cohorts 1 and 2 and 117.9 months for Cohorts 4 and 5. These distributions are characterized by positive skew and large standard deviations, indicating that the distributions are concentrated on the low end of the scale. The mean number of adults in the household ($M = 1.6$ ($s = .8$) for Cohorts 1 and 2, and $M = 1.5$ ($s = .8$) for Cohorts 4 and 5) indicates that a substantial percentage of the households had single parents. Employment rate was extremely low in both groups. Of the respondents, 34% of the heads of household were employed in Cohort 1 and 2, and 23% in Cohorts 4 and 5.

EXPERIMENT 1

Materials and procedure

Measures of language proficiency, intellectual ability, and metalinguistic ability were administered to all subjects. These are described below.

Language proficiency   All subjects were administered tests of receptive vocabulary in both English and Spanish at all four times in order to estimate their proficiency levels in the languages. The Peabody Picture Vocabulary Test (Dunn, 1965) was used as the measure of English proficiency. An adaptation of this test for the Spanish-speaking population in New York, developed by Wiener, Simmond, and Weiss (1978), was used for Spanish. These measures will be abbreviated as EPVT and SPVT, respectively, in the remainder of the paper. The percentile scores provided for raw scores must be interpreted with great caution because the standardization population is different from our subject population. This caution bears particular attention in the case of the Spanish test, for which we had to construct our own estimates based on the authors' reported data (details are reported in Hakuta, 1984). This was necessary because the authors of the Spanish version reported norms based on raw scores where all 150 items on the test are administered, whereas the instruction requires that testing be stopped once the subject makes a critical number of errors.

Two measures were used to validate the results of the EPVT and SPVT. One

Table 2. *Mean proficiency scores and percentile equivalents for English and Spanish in Cohorts 1 and 2*

		Mean EPVT	Percentile	Mean SPVT	Percentile
Cohort 1	Time 1	30.62	2	53.78	60
	Time 2	40.86	3	56.58	55
	Time 3	47.10	16	60.17	60
	Time 4	53.31	6	64.73	67
Cohort 2	Time 1	45.33	0	63.13	60
	Time 2	50.69	2	65.17	70
	Time 3	56.46	4	68.37	65
	Time 4	59.75	8	71.27	65

measure was a story-retelling task: 40 children in Cohorts 1 and 2 were read two stories, one in Spanish and one in English. Each story contained 14 sentences, paired with cartoon depictions of the sentences. Following presentation, the subjects were asked to retell the story. Their performance was rated on a 5-point scale (1 = minimally proficient, 5 = fully proficient). Inter-rater reliability for the English version was .96, and for the Spanish measure, .89. The correlation between the English version of this task and the EPVT was .82; the correlation between the Spanish story-retelling task and the SPVT was .36 (there was less variance in the Spanish score in either of the measures). In addition, the EPVT and SPVT scores were compared to those of the English and Spanish versions of the Language Assessment Scale (De Avila & Duncan, 1981), administered to 49 children in the four cohorts. The correlation between the Spanish tests was .55, and .62 between the English tests.

The raw scores on the EPVT and the SPVT indicated that there was not much overlap between Spanish and English proficiency in either Cohorts 1 or 2. Whereas the majority of Spanish raw scores fell above the 40th percentile at all four testing times, the majority of the English raw scores fell below the 40th percentile. When the scores did overlap, it occurred mostly in the 20th to 60th percentile range. It must be emphasized, however, that all subjects were Spanish dominant; even low scores in Spanish were higher than high scores in English. Thus English proficiency functioned as the measure of bilingualism in the children – the more English they knew, the more bilingual they were. The mean raw EPVT and SPVT scores for each testing time and the corresponding percentiles are presented in Table 2 for Cohorts 1 and 2. These mean raw scores were used for a mean-split classification of the bilingual children into high and low proficiency groups in the two languages.

Intellectual ability Intellectual ability was measured with the Raven Coloured Progressive Matrices (Raven, Court, & Raven, 1976). This measure was administered in the more dominant language, Spanish. The correlation between the raw Raven's score (with age partialled out) and the metalinguistic measures fell between .27 and .37 at the four testing times in both cohorts. The correlations

Table 3. *Examples of items in the*
ungrammaticality task

Noun/article error (gender) (NAR)
*La perro es grande (El perro)
*Un niña es mi amiga (La niña)
*El flor tiené muchos colores (La flor)
*José cortó la arbol (el arbol)

Noun/adjective error (gender) (NAD)
*La casa es pequeño (pequeña)
*El pescado es bien bonita (bonito)

Word order error (WO)
*La casa tiene techo no (no tiene techo)
*Lapiz un dame (Dame un lapiz)
*Mi mama los me compró (me los)
*El está lo vendiendo (lo está)

Noun/verb error (singular/plural) (NV)
*Pablo y José es un primo (son primos)
*El pájaro y el gato comió bien (comieron)
*Las señoras se fué a comer (se fueron, La señora)
*Los animales tomó agua (tomaron, el animal)

Adverb of time/verb tense error (T)
Juan fué a la tienda mañana (irá, ayer/hoy)
Juan irá ayer a la playa (fué, mañana/hoy)
Ayer estoy limpiando el barco (ahora/hoy, estaba)
Mañana comí muchos dulces (ayer/hoy, comeré)

between the Raven's score and the EPVT and SPVT scores varied between .1 and .4 in the two cohorts at the four testing times. Low and high proficiency groups in the two languages, however, did not differ significantly in the Raven's with age partialled out.

Metalinguistic ability A classic means of assessing metalinguistic awareness in preschool and early elementary-school monolingual children has been to ask them to note errors in constructions and to correct them. Most experiments have tested children's ability to note errors that render constructions inconsistent with world knowledge (such as "ride the picture") (de Villiers & de Villiers, 1972; Hakes, 1980; Howe & Hillman, 1973; James & Miller, 1973). Only a few studies have examined children's ability to note errors in constructions that have a plausible interpretation and where the error is more formal in kind (such as "Girl is swimming") (Beilin, 1975; Gleitman, Gleitman, & Shipley, 1972; Ryan & Ledger, 1979, in monolingual children; Galambos & Goldin-Meadow, 1983, in bilingual children). These experiments indicate that monolingual children have a difficult time noting and correcting errors of this kind before the age of 5;6 to 6;0, even though their speech is devoid of such errors, but that bilingual

Table 4. *Examples of types of corrections in the ungrammaticality task*

Ungrammatical construction	Correction
Grammar-oriented correction	
1. *La perro es grande	El perro es grande
2. *El pescado es bien bonita	El pescado es bonito
3. *Juan fué a la tienda mañana	Juan fué a la tienda ayer
Content-oriented correction	
4. *La perro es grande	El perro es pequeño
5. *La perro es grande	Y me va a morder

*Ungrammatical sentences.

children *who are proficient in both languages* can easily note such errors at the age of 4;6. Young monolingual children appear to focus on the message conveyed by constructions, whereas bilingual children readily focus on the form of constructions upon demand.

The constructions in the metalinguistic task in this experiment were patterned after those used in Galambos and Goldin-Meadow (1983). Some of the types of errors that were difficult to note, and some of the types that were easy to note were included in our task. We were interested in testing whether the Galambos and Goldin-Meadow results could be replicated with a substantially different bilingual population. The error types used in our task were the following: noun/article errors involving gender (NAR), noun/adjective errors involving gender (NAD), word-order errors (WO), noun/verb errors involving number (NV), and adverb of time/verb-tense errors (T). The constructions used in the task are listed on Table 3 by type of error.

There were 7 fillers and 18 ungrammatical constructions in the metalinguistic task. Each child was read the constructions individually. He was then asked to judge whether it had been said the right way ("¿así se dice?"), and finally he was asked to correct the error he had noted ("¿como se dice pues?").

Following procedures developed in Galambos and Goldin-Meadow (1983), responses were first scored as "noted" (1) or "not noted" (0). Only constructions that were noted and corrected were coded as "noted." Those that were noted but not corrected were coded as not noted. This was done to eliminate false positives. Responses to fillers were also examined. This enabled us to assess the reliability of the children's performance. Corrections for the errors noted were coded next. The primary decision in coding the corrections was to decide whether the correction indicated that the child had attended to the grammatical properties of the construction or whether it indicated that he had attended to its content. A correction was coded as "grammar-oriented" if a child had corrected the perceived error in the grammar without also making a substantial and unwarranted change in the content of the construction. Examples of this can be found in Table 4. A correction was coded as "content-oriented" if the child had made a substantial change in the propositional content of the construction when this

change was not warranted (e.g., #4- 5 in Table 4). The correction of the error in responses such as #4 "El perro es pequeño" was judged to be an accidental byproduct of the content-oriented correction. Inter-rater agreement for the corrections, based on 100 responses from 25 different subjects, was 95%.

Results

The primary analyses that will be reported here are unweighted means Analyses of Variance with two repeated measures (Time of Testing with four levels and Type of Item with five levels) and two grouping factors (EPVT and SPVT, each with two levels, Low and High). There were two dependent variables analyzed separately: errors noted and type of correction. The score for type of correction was based on the number of errors that were noted, rather than on the total number of items in the task.

Separate analyses were conducted for Cohorts 1 and 2 because an initial analysis that included Cohort as a grouping factor and using number of errors noted as the dependent variable showed significant interactions of Cohort with other variables. The Time × Cohort interaction, $F(3,288) = 4.54$, $p < .05$, indicated that there was greater improvement over time for Cohort 1 than for Cohort 2. This was most likely due to the older cohort approaching a ceiling, particularly for the items that were easier to note (NAR, NAD, and WO). This was apparent in a Type × Time × Cohort interaction, $F(12,1152) = 2.965$, $p < .001$.

In order to be able to include Time as a repeated measures factor in the analyses, a decision had to be made as to how the subjects could be classified with respect to EPVT and SPVT proficiency levels. Because proficiency had been measured at each of the four time periods, there were potentially four different classifications for each subject. Although the correlations of the proficiency measures across time were respectable (.60 to .75 in both Spanish and English), there was a sizeable number of subjects whose proficiency classification changed across time. Rather than taking one of the four proficiency classifications as the representative one, we decided that a more sensible approach for this analysis would be to average the EPVT and SPVT scores for each subject across the four time periods and to divide the children into low and high proficiency groups for English and Spanish at the mean of each distribution.

The cell sizes for the two cohorts were the following: Cohort 1, Low Spanish–Low English, $n = 15$; Low Spanish–High English, $n = 11$; High Spanish–Low English, $n = 8$; High Spanish–High English, $n = 18$. Cohort 2, Low Spanish–Low English, $n = 18$; Low Spanish–High English, $n = 11$; High Spanish–Low English, $n = 5$; High Spanish–High English, $n = 18$. The uneven distribution of subjects in the four cells for both cohorts is a natural consequence of a general positive correlation between Spanish and English. Although unequal cell sizes may have undesirable statistical consequences depending on a number of parameters (Snedecor & Cochran, 1967), this design enabled us to examine the effect of English proficiency on metalinguistic awareness independently from that of Spanish.

Pearson's correlations between EPVT and SPVT scores were positive and

moderately strong across all sampling periods. For Cohort 1, at the four different times, they were .25 ($p < .05$), .32 ($p < .01$), .50 ($p < .001$), and .41 ($p < .001$), respectively. For Cohort 2, again at the four different times, they were .44 ($p < .001$), .54 ($p < .001$), .47 ($p < .001$), and .52 ($p < .001$), respectively. To assess the degree to which we were successful in creating independence between English proficiency and Spanish proficiency, we conducted analyses of variance with English and Spanish proficiency groupings as independent variables and SPVT as the dependent variable. For each cohort, we conducted analyses for each of the four time periods. The following F-values were obtained for the main effect of English proficiency level on SPVT for Cohort 1: Time 1, $F(1,51) < 1$; Time 2, $F(1,51) = 2.81$, n.s.; Time 3, $F(1,51) = 6.08$, $p < .05$; Time 4, $F(1,51) < 1$. For Cohort 2, the following F-values were obtained: Time 1, $F(1,51) < 1$; Time 2, $F(1,51) = 2.48$, n.s.; Time 3, $F(1,51) < 1$; and Time 4. $F(1,51) = 2.43$, n.s. Thus, we concluded that for both Cohorts 1 and 2, the level of English proficiency did not (with one exception) significantly affect the SPVT scores, and thus that the level of English proficiency was in fact a good indicator of degree of bilingualism in our subjects independent of Spanish proficiency.

Cohort 1   The Analysis of Variance using errors noted as the dependent measure showed a significant main effect for Type, $F(4,192) = 69.36$, $p < .001$. Inspection of the means showed that the items could be grouped into two categories of difficulty. The easy items were: NAR ($M = .70$), NAD ($M = .67$), and WO ($M = .60$). The difficult items to note were: NV ($M = .32$) and T ($M = .31$). Further inspection of the individual item means within the item types revealed that the effects were not due to individual outliers.

A significant effect for Time, $F(3,144) = 27.0$, $p < .001$, showed a steady improvement over time (Time 1 = .39, Time 2 = .49, Time 3 = .56, Time 4 = .64). This improvement is even more striking when the performance on the filler items is taken into account. Whereas at Time 1, errors were wrongly noted in 31% of the fillers, at Time 4, this percentage had reduced to 15%. Time did not interact significantly with Type, $F(12,576) = 1.19$, n.s., suggesting that all item types increased similarly.

Not surprisingly, there was a main effect of Spanish proficiency (SPVT) on metalinguistic awareness, $F(1,48) = 8.83$, $p < .005$, Low SPVT = .43, High SPVT = .61. In addition, however, there was a main effect for EPVT, $F(1,48) = 4.20$, $p < .05$, Low EPVT = .46, High EPVT = .58, suggesting that even within the low overall level of bilingualism in our subject population, degree of bilingualism had an effect on metalinguistic awareness.

An even more interesting finding was the significant three-way interaction between EPVT, SPVT, and Type, $F(4,192) = 2.83$, $p < .05$. The means are presented in Table 5. As can be seen, among the Low SPVT group, the largest difference between the Low and High English groups can be found on the easier items (NAD, NAR, and WO). On the other hand, among the High SPVT group, the largest difference between the Low and High English groups can be found on the hardest item (T), although possibly this result is due to a ceiling effect on the easier items. This indicates that the effects of bilingualism may vary depending

Table 5. *English* × *Spanish* × *Type interaction in the ungrammaticality task*

Language group	NAD	NAR	WO	T	NV
Low Spanish					
Low English	.43	.50	.40	.18	.16
High English	.71	.73	.65	.28	.28
High Spanish					
Low English	.81	.77	.62	.30	.43
High English	.74	.82	.74	.49	.40

on the characteristics of the items used to tap metalinguistic awareness and could possibly also vary depending on the level of proficiency in the primary language.

The ANOVA using type of correction (specifically, grammar-oriented correction) as the dependent measure showed similar but less robust results in general. There was a significant effect of Type, $F(4,192) = 19.84$, $p < .001$. As in the analysis using errors noted as the dependent measure, the easy items were NAR ($M = .80$), NAD ($M = .65$), and WO ($M = .70$). The hard items were NV ($M = .56$) and T ($M = .46$). Thus, the items that were the easiest to note were also the easiest to correct in a grammar-oriented way.

There was also an effect of Time, $F(3,144) = 36.43$, $p < .001$, with an increasingly greater proportion of grammar-oriented corrections over time; thus, while at Time 1, 43% of the corrections were grammar-oriented, at Time 4, 80% were grammar-oriented.

The degree of Spanish proficiency was also found to affect the proportion of grammar-oriented corrections offered, $F(1,48) = 6.86$, $p < .01$, Low Spanish $= .54$, High Spanish $= .72$. The effect of EPVT was not significant, although the same tendency emerged as in the previous analysis, $F(1,48) = 2.78$, $p = .102$, Low English $= .58$, High English $= .69$.

Cohort 2 As in Cohort 1, an Analysis of Variance using errors noted as a dependent measure showed a significant main effect for Type, $F(4,192) = 31.78$, $p < .001$. The pattern of item difficulty was the same as that observed in Cohort 1 (NAR $M = .89$, NAD $M = .82$, WO $M = .79$, NV $M = .52$, T $M = .57$). There were no outliers in the individual items.

Although Time showed a significant main effect, the magnitude of the improvement over time was considerably less than that observed for Cohort 1, $F(3,144) = 5.57$, $p < .01$, Time 1 $= .66$, Time 2 $= .70$, Time 3 $= .75$, Time 4 $= .76$. A significant Type × Time interaction, $F(12,576) = 4.49$, $p < .001$, showed that the easier items (NAR, NAD, WO) were approaching ceiling level performance, whereas most of the improvement over time could be attributed to increased performance on the more difficult items (NV and Time). This interaction was not significant in Cohort 1, showing similar improvement across item types.

As in Cohort 1, there was a main effect of Spanish level on the number of errors noted, $F(1,48) = 5.65$, $p < .05$, Low Spanish = .65, High Spanish = .79. Although the main effect for English level was not significant, $F(1,48) = 1.76$, n.s., the interaction between English and Spanish levels was very close to significant, $F(1,48) = 3.97$, $p = .053$. An interesting compensatory relationship between the two languages is suggested by this interaction. The means indicate that the effect of English was stronger at Low Spanish, or, conversely, that the effect of Spanish was stronger at Low English (Low Spanish, Low English = .55; Low Spanish, High English = .74; High Spanish, Low English = .81; High Spanish, High English = .77).

A significant four-way interaction of Spanish × English × Type × Time, $F(12,576) = 2.615$, $p < .01$, provides an instructive example of the complexity of the relationship between degree of bilingualism and metalinguistic awareness. As the two-way interaction suggested, there were no effects of English level within the High level of Spanish. This effect was consistent across Time and Item levels. Moreover, within High Spanish, improvement over time was found on the more difficult items (NV and T), and not on the easier items, which had approached ceiling.

A different pattern emerged within Low Spanish, where effects were observed at different combinations of Time and Item levels. For the easier items (particularly NAR, NAD), the effect of English was strongest at Time 1, and diminished over time. For the difficult items (NV, T), the effect of English was generally constant over time. Thus, as in Cohort 1, the effects of bilingualism differed depending on the time of testing and the items tested.

The ANOVA using the type of correction as the dependent measure showed essentially the same pattern of results, although moderated by the fact that for many of the items, the proportion of grammatically oriented responses was close to ceiling. There was a significant main effect for Type, $F(4,192) = 6.917$, $p < .001$, NAR $M = .95$, NAD $M = .85$, WO $M = .87$, NV $M = .75$, T $M = .77$. Time was also significant, $F(3,144) = 12.116$, $p < .001$, showing a linear increase with time approaching ceiling by Time 4 (Time 1 = .74, Time 2 = .83, Time 3 = .88, Time 4 = .90).

The level of Spanish proficiency also affected the proportion of grammar-oriented responses, $F(1,48) = 9.90$, $p < .01$, Low Spanish = .75, High Spanish = .93. English did not attain significance, $F(1,48) = 2.23$, n.s. A near-significant interaction of English × Spanish, $F(1,48) = 3.78$, $p = .058$, indicated the same pattern as in the first analysis, with the greater effect of English showing up at Low Spanish, and the greater effect of Spanish showing up at Low English.

The Spanish × Type interaction was close to significant, $F(4,192) = 2.367$, $p = .055$. The means indicated varying levels of performance depending on item type among the Low Spanish group (NAR = .90, NAD = .76, WO = .83, NV = .61, T = .64), but consistent near-ceiling performance across item types for the High Spanish group (NAR = 1.0, NAD = .95, WO = .90, NV = .89, T = .90). The Spanish × Time interaction was significant, $F(3,144) = 3.007$, $p < .05$, showing no effect of Spanish at Times 3 and 4, due to the attainment of ceiling by these times.

Discussion

The primary finding in this experiment from a psycholinguistic viewpoint was that the types of items in the metalinguistic task differed in how difficult they were to note and correct. We found that word order errors (WO) and errors involving gender agreement (NAR and NAD) were easy to note and correct. In contrast, errors of temporal agreement between adverb and verb tense (T) and errors involving number agreement between noun and verb (NV) were significantly more difficult to note and correct. These results are consistent with those reported by Galambos and Goldin-Meadow (1986), who used similar items with balanced and highly proficient Spanish-English bilinguals.

As an explanation for the item type differences, Galambos and Goldin-Meadow suggested that factors such as the saliency and complexity of the errors might affect the level of difficulty of the items. They postulated that gender errors were salient due to their high functional load in Spanish. Supporting evidence for this can be found in the fact that gender concordance is acquired very early in Spanish (Fantini, 1974). Also, in their experiment, they found that even their prekindergarten children, who were limited in their knowledge about the rules of grammar, were aware of the rules governing gender. In accounting for the saliency of word-order errors, they noted that children in acquiring their first language are sensitive to the ordering of words (Gleitman & Wanner, 1982).

The lack of saliency of noun-verb number agreement and adverb-verb tense agreement was argued once again on the basis of findings from language acquisition, where it has been reported that markers of tense and number on verbs are acquired relatively late (Berko Gleason, 1985; Dale, 1976). The children tested in the Galambos and Goldin-Meadow study as well as in ours are well past this stage of language acquisition and have mastered these rules of grammar in speech. However, the claim is that the very same factor that influences their lateness of acquisition, namely saliency, also influences the ease of noting errors of these types.

In addition to the psycholinguistic findings above, this study revealed the importance of language proficiency factors in the native and second language on metalinguistic awareness. We found a consistent effect of native language (Spanish) proficiency on the abilities to note and correct errors. In addition, the interaction of Spanish with English indicated that if the level of bilingualism was low, the level of native language proficiency (Spanish) was particularly important, suggesting an interesting compensatory relationship between the two languages.

The effect of bilingualism (English) was found to vary depending on the level of proficiency of the native language (Spanish) and the difficulty of items. At high levels of Spanish, the effect of English mostly occurred on the hardest items in both the noting and correcting measures. In contrast, at low levels of native language proficiency, the effect of bilingualism was more generally evident across item types. This is consistent with Galambos and Goldin-Meadow's data, in which their first-grade highly proficient bilingual subjects particularly outperformed the monolingual control subjects on the hardest items.

We find an information-processing framework to be useful in conceptualizing these results. It would appear that if an error were salient, only superficial

attention to form might be needed to note the error, although the form of the construction would still have to be processed exhaustively in order to be successfully corrected. In contrast, if the error were difficult to note, it would be necessary for a subject to attend closely to the form of the construction and process it exhaustively in order to note and correct the error. Particularly relevant in the information-processing approach is the emphasis on limited processing capacity for tasks. This view would predict that situations of competing demands for attentional resources would result in decreased processing depth. It would also predict that experiences that reduce the load on working memory would permit attention to be more fully devoted to the task at hand (Anderson, 1982; Sternberg, 1984). In the context of our particular experiment, this would mean that a lack of high linguistic proficiency in Spanish would be expected to interfere with the ability to attend to and process the form of constructions exhaustively, thus decreasing the chances of noting and correcting errors effectively. This prediction is consonant with our results, which indicate that the children who were not highly proficient in Spanish were significantly less able to note and correct errors than the children who were highly proficient in Spanish.

With regard to the effects of bilingualism on metalinguistic awareness in our experiment, the information-processing approach successfully accounts for our findings that bilingualism by and large enhances the metalinguistic abilities to note errors and correct errors (see Bialystok & Ryan, 1985, for a similar perspective). The bilingual experience requires that the form of the two languages being learned be attended to on a routine basis. Experience at attending to form would be predicted to facilitate any task that required a child to focus on form upon demand. Our findings indicate that this facility at attending to form in the more highly bilingual children served to compensate generally for the loss of processing efficiency due to a lower level of Spanish proficiency. On the other hand, in our subjects with a high level of Spanish proficiency, who were able to attend to form with greater ease, the bilingual experience (particularly in our youngest group of subjects) served to alleviate especially the processing demands of the more difficult items.

To conclude, we found that the development of the native language to its fullest is beneficial to metalinguistic awareness. In addition, we found that the degree of bilingualism was also important. The extent to which this effect could be observed varied, however, on the level of difficulty of the items. This suggests that the results might have been different had another mix of easy and difficult items been used in measuring metalinguistic awareness. The fact that level of native language proficiency also interacted with degree of bilingualism suggests the importance of closely examining both languages in characterizing bilingual children.

EXPERIMENT 2

Materials and procedure

Measures of language proficiency, intellectual ability, and metalinguistic ability were administered to all subjects. These are described below.

Table 6. *Mean proficiency scores and percentile equivalents*
for English and Spanish in Cohorts 4 and 5

		Mean EPVT	Percentile	Mean SPVT	Percentile
Cohort 4	Time 1	59.08	4	79.61	75
	Time 2	70.07	12	86.70	85
Cohort 5	Time 1	62.78	2	84.80	80
	Time 2	72.64	9	91.49	90

Language proficiency   All subjects were administered at the two testing times specified by the Peabody Picture Vocabulary Test in English (EPVT) to measure English proficiency, and according to the adaptation of this test in Spanish (SPVT) (described above) to measure Spanish proficiency.

As in Experiment 1, the EPVT and SPVT raw scores suggested little overlap between Spanish and English proficiency in either Cohort 4 or Cohort 5. Whereas in both cohorts the EPVT scores at Time 1 fell mostly below the 24th percentile, and at Time 2, below the 50th percentile, SPVT scores at Time 1 and Time 2 fell mostly above the 50th percentile score. Thus, these children were "unbalanced" bilinguals – quite proficient in Spanish while of low proficiency in English. As in Experiment 1, the more English the children knew, the more bilingual they were. Therefore the EPVT score was used in the analyses as the measure of bilingualism in the children.

The mean raw EPVT and SPVT scores for each testing time are presented in Table 6 for Cohorts 4 and 5. These scores were used to classify the children into high and low proficiency groups in the two languages.

Intellectual ability   Intellectual ability was measured with the Raven Coloured Progressive Matrices. As in Experiment 1, this measure was administered in the more dominant language, Spanish. Semi-partial correlations between the Raven score (with age partialled out) and the metalinguistic measures at the two testing times and for the two cohorts ranged from .07 to .25. The correlations between proficiency scores and Raven scores were lower than .20. Moreover, no significant differences were found in Raven scores for low and high proficiency groups in either Spanish or English.

Metalinguistic ability   Metalinguistic ability was tested, at both times, with an ambiguity task. We chose this task because the detection of ambiguity, unlike the detection of ungrammaticality, has been found to be a late-developing metalinguistic ability. For instance, Shultz and Pilon (1973) found that ambiguity in homophonous items (pears/pairs) was easily detected and paraphrased by fourth graders, but that the ambiguity in polysemous items was not. The appreciation of ambiguity in jokes was tested by Hirsh-Pasek, Gleitman, and Gleitman (1978), who found that children in grades 3 to 6 had a difficult time recognizing phonetic ambiguity at morpheme boundaries (engineer/engine ear) as well as explaining surface structure (man-eating fish/man eating fish) and deep structure (grand-

Table 7. *Examples of items in the ambiguity task*

Polysemous	
Elena vió las cartas de María	Elena saw Maria's letters/cards
La pluma verde está en la mesa	The green pen/feather is on the table
El obrero pintó el banco de gris	The workman painted the bank/bench gray
En California se ven muchas estrellas	In California many stars (actors/stars) can be seen
El músico tocó la guitarra	The musician played/touched the guitar
Homophonous	
Cuando Claudio entró, Juan se cayó/ calló	When Claudio entered, Juan fell/ became quiet
Mis dos amigos se fueron a casar/cazar	My two friends went to get married/ went to hunt
Luis botó/votó y luego recogió la ropa	Luis dropped/voted and then picked up the clothing
Phonetically ambiguous (at morphological boundaries)	
René se escondió del agente/de la gente	Rene hid from the policeman/the people
El loro/el oro está en la cueva	The parrot/gold is in the cave
José separó/se paró y limpió los juguetes	Jose separated/stood up and cleaned the toys
Elena va a pagar/apagar las velas	Elena is going to pay for/extinguish the candles
La costurera va a cortar/acortar la falda	The seamstress is going to cut/shorten the skirt

mother for Thanksgiving: to eat or to eat with) ambiguity. Unlike Shultz and Pilon (1973), however, they found polysemy (bark/bark) to be easily detected by children in grades 1 to 3.

There were 19 items in our metalinguistic task. These are listed in Table 7. As indicated in the table, there were five polysemous items, three homophonous items, and five phonetically ambiguous items that could be segmented in two ways to derive different interpretations. There were also three filler items, that is, constructions with only one meaning. These items were recorded in a randomized order by a native speaker of Spanish.

After hearing each of the taped sentences, the children were asked to say how many meanings the sentence had ("¿Entendiste esta oración de una manera o de dos maneras?"). Next, the children were asked to paraphrase each of the meanings identified (for instance, a child who identified two meanings was asked "Explícame la primera manera como tu entendiste la oración," followed by "Explícame la segunda manera como tu entendiste la oración"). Finally, the children were shown two pictures depicting each of the two possible meanings, and were asked if they recognized both meanings (while pointing to each picture, the child was asked, "¿La entendiste de esta manera?"). For the filler items, the second picture did not depict the meaning of the sentence.

The responses were scored for each of the three probes listed above. For the

first probe (CLAIM), each response was scored as 0, 1, or 2, to denote the number of meanings identified by the subject. For the second probe (PRO-DUCE), each response was also scored as 0, 1, or 2, depending on the number of meanings correctly paraphrased by the child. For the third probe (RECOG-NIZE), the number of pictures positively identified was scored as 0, 1, or 2. For each of these measures, a mean score was derived for each subject on each of the item types discussed above (Polysemous, Homophonous, and Phonetically Ambiguous).

Results

Of the three dependent measures mentioned above, an analysis of the responses on the filler items revealed that PRODUCE was the most reliable measure, followed by CLAIM and by RECOGNIZE. In both grades, a substantial propor-tion of false alarms was found on the RECOGNITION measure. Across Time and Cohort variables, the mean number of meanings recognized on the filler items was 1.3, when ideally it should have been 1.0. For CLAIM, the mean on the filler items was 1.1. On PRODUCE, it was 1.0. Based on these data, the principal analyses were conducted on the PRODUCE measure, with supplemen-tary analyses on CLAIM.

An overall analysis with Cohort as a factor revealed that there were no signifi-cant differences between Cohorts 4 and 5 on the three dependent measures, nor were there differences when they were compared on each of the item types. Thus, the analysis that will be reported here combines both cohorts.

However, the proficiency classification of the subjects was conducted sepa-rately for each cohort to avoid a confound between age and proficiency group-ings. Mean EPVT and SPVT scores were calculated and used as the division point for grouping subjects into low and high proficiency groupings in the two languages. The cell sizes differed slightly depending on whether the groups were divided according to Time 1 or Time 2 EPVT and SPVT scores. For Time 1, the following were the cell sizes: Low Spanish–Low English, $n = 39$; Low Spanish–High English, $n = 18$; High Spanish–Low English, $n = 16$; and High Spanish–High English, $n = 34$. For Time 2, Low Spanish–Low English, $n = 38$; Low Spanish–High English, $n = 16$; High Spanish–Low English, $n = 22$; High Spanish–High English, $n = 31$.

The results will be reported first for each time of testing separately. As in the study for Cohort 1 and 2, an unweighted means Analysis of Variance was used. EPVT (Low and High) and SPVT (Low and High) were used as grouping factors, and the three types of items discussed above (Polysemous/ Homo-phonous/ Phonetically Ambiguous) were used as the three levels of the repeated measures factor.

In the longitudinal analysis, Time (with two levels separated by a year) was included as a second repeated measures factor. Unlike our analysis for Cohort 1 and 2, we chose to conduct two separate analyses involving Time. In the first analysis, we classified subjects into proficiency groups based on their EPVT and SPVT scores at Time 1, and in the second analysis, based on their scores at Time

2. We reasoned that this approach was preferable to averaging the proficiency scores at the two testing periods. Furthermore, unlike the method for Cohorts 1 and 2, this was feasible because there were only two testing times.

Finally, Pearson's correlations between EPVT and SPVT were found to be positive and moderately strong for the two sampling periods: at Time 1, $r = .36$ ($p < .001$), and at Time 2, $r = .46$ ($p < .001$). In order to check on how successful we were in our attempt to create independence between English and Spanish proficiency through our ANOVA design, we conducted parallel analyses as those used for Cohorts 1 and 2. Thus, English and Spanish proficiency grouping factors were used in an ANOVA with SPVT as the dependent variable. When the Time 1 groupings were used, the main effect of English proficiency grouping was not significant for SPVT scores at Time 1, $F(1,103) < 1$, n.s., nor at Time 2, $F(1,103) = 1.00$, n.s. When Time 2 groupings were used, English grouping did have an effect on SPVT scores. For SPVT at Time 1, $F(1,103) = 15.63$, $p < .001$, and for SPVT at Time 2, $F(1,103) = 8.92$, $p < .01$. This suggested that our analytic strategy of forcing independence between English and Spanish was not as successful as it was for the first study with younger subjects. Thus, any effects related to the English variable at Time 2 cannot be totally attributed to bilingualism, as it is to some measure confounded with Spanish proficiency. We will discuss the implications of this in our discussion.

Time 1 analysis   The Analysis of Variance for Time 1 using PRODUCE as the dependent measure showed a significant main effect for item type, $F(2,206) = 74.28$, $p < .001$, Phon = 1.01, Homo = 1.17, Poly = 1.42. Although we expected phonetic ambiguities to be more difficult to detect, we were surprised by the difference between homophonous and polysemous items, as this implies that orthography plays a role in the processing of these items.

There was also a main effect for Spanish, $F(1,103) = 6.76$, $p < .05$, Low Spanish = 1.13, High Spanish = 1.27. The effect for English was not significant.

Time 2 analysis   The main effect for item type was more robust than for Time 1, $F(2,206) = 106.06$, $p < .001$, and followed the same pattern (Phon = 1.19, Homo = 1.35, Poly = 1.65).

The main effect for Spanish was again significant, $F(1,103) = 20.88$, $p < .001$, Low Spanish = 1.30, High Spanish = 1.50. Moreover, there was also a main effect for English, $F(1,103) = 4.38$, $p < .05$, Low English = 1.35, High English = 1.44. Although this could suggest that bilingualism was related to metalinguistic awareness at this period, a firm conclusion cannot be advanced due to the fact that independence between English and Spanish proficiency could not be ascertained at Time 2.

Interactions of Type with both English, $F(2,206) = 2.22$, $p = .11$, and with Spanish, $F(2,206) = 2.23$, $p = .11$, were close to significant, and indicated patterns of potential interest. The effect of English apparently was confined to the phonetically ambiguous and the polysemous items, whereas the greatest effect of Spanish occurred on the homophonous items.

Longitudinal analysis   The results of the analysis where subjects were classified by Time 1 proficiency levels replicated the main effects of Type and of SPVT found in the separate analyses reported above. In addition, however, there was a significant improvement over Time, $F(1,103) = 74.48$, $p < .001$, Time 1 = 1.20, Time 2 = 1.40, and an interaction of English by Type by Time, $F(2,206)$ = 3.83, $p < .05$. Examination of the means showed that this interaction was attributable to the fact that the largest improvement over time was found among the High English group on the Phonetically Ambiguous type of items (which was also the hardest type).

The results of the analysis using Time 2 proficiency classifications yielded essentially the same results, with the exception that there was also a main effect of English, $F(1,103) = 3.81$, $p = .05$. This main effect of English was also found in the separate analysis conducted for Time 2 reported above. The limitations of interpreting this effect of English as an effect of bilingualism was discussed above.

Analyses using CLAIM as the dependent measure showed the same pattern of results, and therefore need no further discussion. However, the fact that the results for CLAIM and PRODUCE were similar suggests that CLAIM, although somewhat less reliable than PRODUCE, adequately measures the ability to detect ambiguity in language.

Discussion

As in Experiment 1, we found that the different types of items varied in their level of difficulty. Phonetic ambiguity was the hardest to detect and paraphrase. Homophonous items were of intermediate difficulty, whereas polysemous items were substantially easier than the other types. Our results replicate the findings of Hirsh-Pasek, Gleitman, and Gleitman (1978) with monolingual English-speaking children. They found that the recognition of phonetic ambiguity in jokes was more difficult than the recognition of ambiguity in jokes using polysemy. Our results, however, are not consistent with those of Shultz and Pilon (1973), who found that homophony was easier to detect than polysemy. These item differences point to the importance of distinguishing different types of ambiguity in studying metalinguistic awareness. Shultz and Pilon, for example, did not distinguish homophonous and phonetically ambiguous items in their study.

A possible explanation for the difference in difficulty of the items observed in our experiment can also be found in the information-processing literature. From this perspective, a prediction could be made that, to detect phonetic ambiguity, it would be necessary to encode the construction quite exhaustively in order to be able to restructure the information. A fair bit of cognitive control would be required to accomplish this. Detecting polysemy, on the other hand, would be much easier. Only a small amount of cognitive control would be required to reinterpret the meaning of a word, as automatized procedures would probably already have been developed to access familiar meanings of a word. Finally, the link between the meanings of homophonous items would probably be less strong and less automatized than that for polysemous items, thus making the ambiguity harder to detect. It would still, however, be easier to detect than that for phonet-

ically ambiguous items, as the phonetic stream would not have to be parsed in two different ways.

Language proficiency variables were also found to be of relevance to metalinguistic awareness in this experiment. The effect of native language proficiency (Spanish) was robust across metalinguistic measures and times of testing. As in Experiment 1, this effect of native-language proficiency on metalinguistic awareness could be explained through an information-processing perspective whereby linguistic fluency would maximize the attentional resources available for the task. In addition, the interaction of Spanish proficiency with item type showed that the effect of Spanish was greatest for homophonous items. A possible interpretation of this result is that with increasing proficiency, the links between orthographically dissimilar but phonetically similar words get strengthened and the related meanings become easier to access. Possibly, orthographic coding becomes secondary with increasing proficiency, and such items become more like polysemous items.

An indication for an effect of bilingualism (English) on metalinguistic awareness was only found at Time 2. It should be recalled that this was also the time period when we could not establish independence between Spanish and English proficiency. This unfortunate confound should perhaps be thought of as a reflection of the realities in the bilingual population. As students increasingly gain proficiency in English, it is likely that the proficiencies in the two languages will become more correlated, as Cummins argued in proposing his theory of linguistic interdependence (Cummins, 1984). Thus, with older bilingual subjects, it is likely that independence of the two languages would be difficult to establish.

If indeed the increased metalinguistic awareness could be attributed to the bilingualism of these subjects, however, there are several possible explanations for why it appears at Time 2, and not Time 1. At Time 2, the subjects were on average more proficient in English than at Time 1. It is conceivable that there is a "threshold effect" (Cummins, 1976) whereby a certain level of proficiency in a second language must be attained before a positive effect of bilingualism on metalinguistic awareness takes place. To rephrase this within an information-processing framework, a critical level of practice in evaluating the formal properties of language may be necessary before expertise is developed. Without such expertise, exhaustive processing of information would not be facilitated. Alternatively, it could be that the metalinguistic task used here was more sensitive in detecting variability due to bilingualism at an older age.

The longitudinal analysis indicated that the detection of phonetic ambiguity (the most difficult type) improved most dramatically for the High English group. This apparent interaction of English proficiency with item type once again attests to the complexity of the effects of bilingualism on metalinguistic awareness.

In sum, although the metalinguistic task in Experiment 2 was different from that used in Experiment 1 and the subjects were of different age, a similar set of conclusions can be drawn. First, the selection of items used in a task can strongly influence the pattern of results. Second, a high level of proficiency in the native tongue is beneficial to metalinguistic awareness. Finally, effects of bilingualism on metalinguistic awareness may vary depending on subject-specific factors,

such as the level of proficiency in the second language, and task-specific factors, such as the types of items in a task. Both of these dimensions must be carefully considered in studying the effects of bilingualism on metalinguistic awareness.

CONCLUSION

In sum, the concern of our study was two-fold. Our first concern was to test the hypothesis that the bilingual experience enhances metalinguistic awareness. Our theoretical framework was that becoming a bilingual requires attention to the form of the languages being learned on a routine basis. Thus, bilingual subjects are expected to do well on tasks requiring such attention to form; within the design of our current study, subjects who were more bilingual were expected to do better than those who were at a low level of bilingual proficiency, given the same level of Spanish proficiency. By and large, our results were consonant with these predictions, particularly in the younger cohorts.

Our second concern in this paper was methodological. The concepts of "bilingualism" and "metalinguistic awareness" have all too commonly been treated as unitary concepts. In this study, we attempted to unpack these concepts by exploring in detail the effects of subject-specific factors (i.e., degree of bilingualism) and task-specific factors (i.e., the items used in the task) on metalinguistic awareness.

Regarding subject-specific factors, we argued for the importance of thinking about bilingualism in terms of the proficiencies of both the native language and the second language. Furthermore, our methodology enabled us to examine the independent contributions as well as the interaction of these factors in accounting for metalinguistic awareness. Presumably, other subject-specific variables not considered in this study, such as sociolinguistic factors related to the bilingual experience (Lambert, 1975), may play an important role in affecting the development of both metalinguistic and other cognitive/linguistic abilities.

Addressing task-specific factors, we highlighted the importance of exploring the psycholinguistic nature of metalinguistic skills. In order to tap metalinguistic skills, it is important to use a variety of tasks as well as to vary the items within each type of task. The pattern of performance, and the interaction of these task-specific variables with subject-specific variables, should reveal the nature of the effect of bilingualism on metalinguistic awareness.

REFERENCES

Anderson, J. R. (1982). Acquisition of cognitive skills. *Psychological Review, 89,* 369–406.

Beilin, H. (1975). *Studies in the cognitive basis of language acquisition.* New York: Academic.

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

Ben-Zeev, S. (1977b). The effects of bilingualism in children from Spanish-English low economic neighborhoods on cognitive development and cognitive strategy. *Working Papers on Bilingualism, 14,* 83–122.

Berko Gleason, J. (1985). *The development of language.* Columbus, OH: Merrill.

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

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

Cummins, J. (1978). Metalinguistic development of children in bilingual education programs: Data from Irish and Canadian Ukranian-English programs. In M. Paradis (Ed.), *The Fourth Lacus Forum* (pp. 29–40). Columbia, SC: Hornbeam.

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

Dale, P. S. (1976). *Language Development.* New York: Holt, Rinehart, & Winston.

De Avila, E. A., & Duncan, S. E. (1981). Bilingualism and cognition: Some recent findings. *NABE Journal, 4,* 15–50.

de Villiers, P. A., & de Villiers, J. G. (1972). Early judgments of semantic and syntactic acceptability by children. *Journal of Psycholinguistic Research, 1,* 299–310.

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

Fantini, A. E. (1974). *Language acquisition of a bilingual child.* Brattleboro, VT: Experimental.

Galambos, S. J., & Goldin-Meadow, S. (1983). Learning a second language and metalinguistic awareness. In A. Chukerman (Ed.), *Papers from the 19th regional meeting of the Chicago Linguistic Society* (pp. 117–133). Chicago: Chicago Linguistic Society.

Galambos, S. J., & Goldin-Meadow, S. (1986). *A study of the development of metalinguistic awareness in bilingual and monolingual children.* Unpublished manuscript, Department of Psychology, Yale University, New Haven.

Gleitman, L. R., Gleitman, H., & Shipley, E. F. (1972). The emergence of the child as grammarian. *Cognition, 1,* 137–164.

Gleitman, L., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–50). New York: Cambridge University Press.

Hakes, D. T. (1980). *The development of metalinguistic abilities in children.* New York: Springer-Verlag.

Hakuta, K. (1984). *The causal relationship between the development of bilingualism, cognitive flexibility, and social-cognitive skills in Hispanic elementary school children* (Final Report No. NIE-G-81-0123). National Institute of Education.

Hirsh-Pasek, K., Gleitman, L. R., & Gleitman, H. (1978). What did the brain say to the mind? A study of the detection and report of ambiguity by young children. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language* (pp. 97–132). New York: Springer-Verlag.

Howe, H. E., & Hillman, D. (1973). The acquisition of semantic restrictions in children. *Journal of Verbal Learning and Verbal Behavior, 12,* 132–139.

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

James, S. L., & Miller, J. F. (1973). Children's awareness of semantic constraints in sentences. *Child Development, 44,* 69–76.

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

Raven, J. C., Court, J. H. & Raven, J. (1976). *Raven's progressive matrices test.* London: H. K. Lewis.

Rosenbaum, T., & Pinker, S. (1983). Word magic revisited: Monolingual and bilingual children's understanding of the word-object relationship. *Child Development, 54,* 773–780.

Ryan, E. B., & Ledger, G. W. (1979). Grammaticality judgments, sentence repetitions, and sentence corrections by children learning to read. *International Journal of Psycholinguistics, 6,* 23–40.

Shultz, T. R., & Pilon, R. (1973). Development of the ability to detect linguistic ambiguity. *Child Development, 44,* 728–733.

Snedecor, G. W., & Cochran, W. G. (1967). *Statistical methods* (6th ed.). Ames, IA: Iowa State University Press.

Sternberg, R. J. (1984). A theory of knowledge acquisition in the development of verbal concepts. *Developmental Review, 4,* 113–138.

Wiener, F. D., Simmond, A. J., & Weiss, F. L. (1978). *Spanish picture vocabulary test.* New York: Marymount Manhattan College.