Measuring Bilingual Children's Receptive Vocabularies

Vivian M. Umbel, Barbara Z. Pearson, María C. Fernández, and D. K. Oller

University of Miami

UMBEL, VIVIAN M.; PEARSON, BARBARA Z.; FERNÁNDEZ, MARÍA C.; and OLLER, D. K. *Measuring Bilingual Children's Receptive Vocabularies.* CHILD DEVELOPMENT, 1992, 63, 1012–1020. Receptive vocabulary of Hispanic children in Miami was tested in both English and Spanish with complementary standardized tests, the Peabody Picture Vocabulary Test (PPVT-R) and the Test de Vocabulario en Imágenes Peabody (TVIP-H). 105 bilingual first graders, of middle to high socioeconomic status relative to national norms, were divided according to the language(s) spoken in their homes. Both groups, whether they spoke only Spanish in the home (OSH) or both English and Spanish in the home (ESH), performed near the mean of 100 in Spanish receptive vocabulary (TVIP-H means 97.0 and 96.5); in contrast, ESH group children scored more than 1 SD higher in English than OSH group children (PPVT-R means 88.0 and 69.7, respectively). It appears, therefore, that learning 2 languages at once does not harm receptive language development in the language of origin, while it does lay the groundwork for superior performance in the majority language. Furthermore, an analysis of translation equivalents, items shared by both tests, shows that a statistically significant portion of bilingual children's lexical knowledge does not overlap in their 2 languages and is therefore not reflected in single-language scores.

In many parts of the world, bilingualism is the norm—a reflection of the need for individuals to participate in more than one language community. In a context of communicative need, it makes little sense to ask whether bilingualism is "better" than monolingualism or vice versa, as speaking only one language is not a realistic option. For many people, to speak only one language would mean exclusion from a meaningful portion of their life.

Nonetheless, in certain educational contexts it makes sense to question whether people can learn two languages as fully as they generally learn one. Can individuals who regularly use two languages achieve a level of proficiency in at least one of them that is equivalent to the norms for that language? Or do we sacrifice a degree of fluency in one language to allow for the other? Is it better for the child to learn one language well before taking on a second; or, if circumstances permit, should the child be exposed to two languages in the early stages of acquisition?

The literature on bilingual language development gives partial and contradictory answers to these questions. On the one hand, the Canadian experience shows bilingual students outperforming monolingual peers on various language measures (Cummins, 1978; Hakuta, 1986; Lambert & Tucker, 1972; Peal & Lambert, 1962), while reports from Scandinavia (Skutnabb-Kangas & Toukamaa, 1976) show that bilinguals, especially those who migrate at the time they start schooling in their first language, may fail to acquire full fluency in either language.

Many factors may give rise to the contradictions between studies. One that is particularly salient is the degree of exposure to each language. Although language experience is considered an important factor in bilingual language development (Ben Zeev, 1977a; Rosenblum & Pinker, 1983), children with various degrees of exposure to both languages are labeled "bilingual." Some authors, for example, count as bilingual anyone with a minority language as a mother tongue, regardless of exposure to or ability in the second language.

The sociocultural context in which bilingual experience occurs may also contribute to the inconsistency of results across bilingual acquisition studies. Lambert (1977)

This research was supported by grant 5 R01 DC00484, NIH NIDCD. Requests for reprints should be sent to Vivian M. Umbel, Bilingualism Study Group, Mailman Center for Child Development, P.O. Box 016820, Miami, FL 33101.

[*Child Development*, 1992, 63, 1012–1020. 0009-3920/92/6304-0018$01.00]

distinguishes between "additive" and "subtractive" forms of bilinguality. In the additive form, both the community and the family attribute positive values to the two languages, and the learning of a second language in no way threatens to replace the first. Subtractive bilinguality develops when one of the languages is not valued outside of the home. Under these circumstances, the language of the economically and culturally more prestigious group tends to replace the minority language. Studies of children experiencing additive bilinguality are more likely to show a bilingual advantage, while those of children experiencing subtractive bilinguality are likely to show a bilingual disadvantage (Sattler & Altes, 1984). In order to provide a clear assessment of bilingual knowledge, it is critical that investigators exercise some control over degree of exposure to the two languages.

Additionally, bilingual acquisition poses a particular problem of measurement from both practical and theoretical perspectives. From a practical standpoint, measurement instruments developed for one language translate inconsistently to another, and it is rare to find an instrument that can be used equally well in two languages (Wilen & Sweeting, 1986). Results from nonstandardized tests are difficult to interpret, and few standardized complementary measures are available (Figueroa, 1990). The APA guidelines for testing linguistic minorities (1985) speak strongly against almost all current practices in this field.

From a theoretical perspective, monolingual norms may be inherently unfair to bilingual children. Standard accounts of language development built on a monolingual model fail to give credit for what Swain (1972) calls the child's total conceptual inventory. Concepts known in one language but not the other are never summed, a policy that potentially produces a serious underestimation of bilinguals' skill. While models of bilingual language organization (Hamers & Blanc, 1989; Paivio & Desrochers, 1980) remain too speculative to support specific psychometric reforms, available tests can themselves demonstrate knowledge their scores fail to quantify.

One area of language development where cross-linguistic standardization has recently been achieved to some extent is vocabulary. The Peabody Picture Vocabulary Test (PPVT) (Dunn, 1965) has been revised and restandardized as the PPVT-R (Dunn & Dunn, 1981). Independent studies have established the concurrent validity and reliability of the PPVT-R with the Language Assessment Scales (Hakuta, 1987) and other language measures (Beck, Black, & Doles, 1985; D'Amato, Gray, & Dean, 1987). A Spanish adaptation of the PPVT-R, the Test de Vocabulario en Imágenes Peabody—Adaptación Hispanoamericana (TVIP-H), became available in 1986 (Dunn, Padilla, Lugo, & Dunn, 1986). Although the test is shorter than its English counterpart, and therefore a little less sensitive, it is a carefully constructed instrument standardized on a large number of children.

The existence of complementary standardized vocabulary tests, one a version of the other, as is the case with the PPVT-R and the TVIP-H, affords new investigative opportunities. On the one hand, it is possible to evaluate vocabulary knowledge of children in both of their languages. Previous research on vocabulary using the PPVT (Argulewicz & Abel, 1984; Ben-Zeev, 1977a, 1977b; Doyle, Champagne, & Segalowitz, 1978; Rosenblum & Pinker, 1983; Serapiglia, 1978) shows young bilingual children scoring in the low average range in English, their second and generally weaker language. However, homemade test translations with no age norms or with only informal "partial standardizations" (Ben Zeev, 1977a, 1977b; Doyle et al., 1978) prevented adequate assessment of the bilingual children's ability in their first language.

Additionally, the two related tests make possible the evaluation of vocabulary items the two tests have in common—items that can be considered translation equivalents of each other. Paired test items permit the discovery of "singlets," words known in one language but not the other. To the extent that knowledge of translation equivalents is restricted to only one language, the monolingual score in one or both languages can be seen to underrepresent the bilingual child's "total conceptual inventory."

Doyle et al. (1978) attempted the assessment of a total vocabulary by summing the scores from two monolingual tests. However, this procedure does not account for the *disjunction* between the single language vocabularies. The greater the disjunction for an individual child, the greater the underrepresentation of his or her lexicalized concepts by any single-language vocabulary score. It is important to determine empirically whether the disjunction is of sufficient

magnitude to significantly affect the proper interpretation of standardized test results.

Goals of the Study

The present work is an attempt to further our understanding of linguistic knowledge in bilingual children by (*a*) evaluating both the English and the Spanish vocabulary knowledge of Hispanic children using standardized tests in each, (*b*) assessing the role of language exposure in vocabulary acquisition by comparing children whose home environments differ in the extent of English use, and (*c*) probing the possibility that bilingual vocabulary knowledge includes an additional dimension compared to monolingual vocabulary, owing in part to the existence of items lexicalized in only one of the two languages.

These goals are pursued in a metropolitan setting unique in the United States. The Hispanic children who are the subjects are of middle to high socioeconomic status relative to national norms. Thus, the study offers a rare opportunity to evaluate bilingual linguistic skills with children who may be expected to be similar in environmental support and linguistic talent to the monolingual comparison groups on which the tests were standardized.

Method

Subjects

One hundred-five English-Spanish bilingual first graders were recruited from four Dade County public schools in predominantly middle-class neighborhoods. The children were divided into two sample groups: first graders from families speaking only Spanish at home (OSH, $n = 31$) and first graders from families speaking both English and Spanish at home (ESH, $n = 74$). Fifty-one percent of the ESH group children had been exposed to both languages since birth; only 14% had parents who had lived in the United States fewer than 12 years. Approximately half of both the OSH and ESH groups consisted of children of Cuban heritage. Other nationalities represented were Puerto Rican, Honduran, Chilean, Venezuelan, Costa Rican, Nicaraguan, Colombian, Argentinean, Peruvian, Mexican, Panamanian, and Guatemalan.

The subjects ranged in age from 5 years 11 months to 8 years 6 months; the average age was 6 years 7 months for both the OSH group and the ESH group. None of the children was repeating the first grade. There were 45 males and 60 females, distributed almost equally across the two language groups. The OSH group was 45% male and the ESH group 42% male.

Procedure

All first graders were given a questionnaire printed in both Spanish and English to be completed by their parents. The questionnaire provided information regarding language(s) spoken at home and SES. Participation exceeded 90% of the students satisfying all the requirements of the study. If the parent(s) consented to having their child participate in the study, and Spanish was spoken in the home, the child was tested individually in a quiet room at the school on the PPVT-R and the TVIP-H (order of presentation balanced across subjects). Both tests were administered by the same bilingual examiner in a single testing session lasting about 20 min.

Tests/Measures Concerning Vocabulary Knowledge

Peabody Picture Vocabulary Test—Revised (Dunn & Dunn, 1981).—The PPVT-R, normed on American English speakers, is a multiple-choice test designed to evaluate receptive vocabulary knowledge. The PPVT-R consists of 175 plates, each containing four pictures. The child is asked to point to the picture that corresponds to the word pronounced by the examiner. Items are arranged in increasing levels of difficulty.

Test de Vocabulario en Imágenes Peabody—Adaptación Hispanoamericana (Dunn et al., 1986).—The TVIP-H is a Spanish adaptation of the PPVT-R, normed on Spanish speakers in Puerto Rico and Mexico. Dunn and his associates attempted to make the test as universal as possible for the different groups that comprise what is considered "Hispanic." The TVIP-H has 125 plates arranged in increasing levels of difficulty.

For both the PPVT-R and the TVIP-H, raw scores are converted into standard scores, with a mean of 100 and a standard deviation of 15.

Translation equivalent (TE) analysis.—Since the TVIP-H was derived from the PPVT-R, the two tests have many items in common. Sixty-one words appear on both forms. These items can be considered translation equivalents of each other. It is possible to get translation equivalent words right in both English and Spanish, wrong in both languages, right in Spanish only, or right in

English only. These last two represent "singlet" vocabulary—words known in one language but not the other.

Translation equivalent presentations for the total sample numbered 1-176, 285 for the OSH group and 891 for the ESH group. The TE presentations per child ranged from 3 to 40 (M = 11.09, SD = 5.35). Only 9% of the children had fewer than six. Analyses of TEs were limited to the 91% of the sample having six or more TE opportunities, 25 OSH and 70 ESH children. Four percentage scores were computed for each child: the percent right in both English and Spanish (PRES), wrong in both English and Spanish (PWES), right in Spanish only (PRSO), and right in English only (PREO). These percentages were derived by totaling the number of TE items the child got right or wrong in each of the categories and dividing by the total number of TE presentations for that child.

Socioeconomic Status

Families were classified according to the occupation of the major wage earner, using the Occupational Levels (OL) set out by Dunn in the PPVT-R standardization (see Table 1). Occupational Levels I through IV were represented among both OSH and ESH parents, but OSH parents as a group held more of the lower-status occupations than the standardization sample for the English Peabody, while the ESH group parents held more of the higher-status occupations than the standardizing sample. The difference in OL between OSH and ESH groups was significant, $F(1, 103) = 15.02$, $p < .001$, the coefficient of association being .36. Therefore, this factor was controlled for in the subsequent analyses of test scores and translation equivalents.

As various occupations are differently stratified by the PPVT-R and the TVIP-H, it is impossible to compare the OSH and ESH samples to the TVIP-H normalization samples without first reclassifying them according to TVIP-H occupational levels. This recategorization changed the classification of 53% of the children in our samples. For example, children of technical and kindred workers classified as Level I according to PPVT-R guidelines received Level II or Level III classifications when categorized by TVIP-H norms. Based on this reclassification, the OSH group sample was roughly comparable to the Spanish Peabody norming sample, while the parents of ESH group children were of generally higher status than the TVIP-H standardization sample.

TABLE 1

A. Occupational Levels (in Percent) of the Only Spanish at Home (OSH) and English and Spanish at Home (ESH) Group Children Compared to the Normalization Sample for the PPVT-R

	Normalization Sample	OSH Group (N = 31)	ESH Group (N = 74)
Level I[a]	25	19	36
Level II	23	16	18
Level III	14	32	12
Level IV	29	32	12
Level V	7	0	0

B. Occupational Levels (in Percent) of the Only Spanish at Home (OSH) and English and Spanish at Home (ESH) Group Children Compared to the Normalization Sample for the TVIP-H

	Normalization Sample	OSH Group (N = 31)	ESH Group (N = 74)
Level I[b]			
Level II	19	22	43
Level III	48	52	51
Level IV	17	26	5
Level V	15	0	0

[a] Level I: professionals, technical, and kindred workers, plus managers and administrators; Level II: sales workers, plus clerical and kindred workers; Level III: craftsmen, foremen, and kindred workers; Level IV: operatives (including transport) plus service workers (including private household); Level V: laborers, farm laborers, or farm foremen.

[b] Level I: high-ranking executives, large-scale proprietors, and major professionals; Level II: business managers, medium-scale proprietors, and lesser professionals; Level III: administrators, small-scale proprietors, semiprofessionals, farm owners, clerical and sales workers, skilled manual workers, machine operators, and housekeepers; Level IV: unskilled workers and farmers; Level V: unemployed. Levels I and II were collapsed by Dunn et al. (1986) to match 1980 U.S. census.

Results

PPVT-R and TVIP-H Performance

PPVT-R and TVIP-H scores were examined to determine any significant effects of home language experience (OSH or ESH) on the children's receptive vocabulary performance (see Table 2). Given the differences found in OL between the OSH and ESH groups, it was necessary to assess the

TABLE 2

PPVT-R AND TVIP-H SCORES OF ONLY SPANISH AT HOME (OSH) AND ENGLISH AND SPANISH AT HOME (ESH) CHILDREN

	OSH	ESH	F	df
PPVT-R:				
Mean (original)	66.94 (17.92)	89.23 (18.33)		
Mean (adjusted)	69.66	88.09	20.62*	1,103
TVIP-H:				
Mean	96.97 (9.30)	96.49 (15.54)	< 1	1,104

NOTE.—Standard deviations are in parentheses.
* $p < .001$.

possible effects of OL on the children's vocabulary scores as a preliminary to these analyses.

Pearson product-moment correlation coefficients were computed to determine the relations between OL and PPVT-R and TVIP-H performance. OL was found to be significantly related to PPVT-R performance, $r(105) = .38$, $p < .001$, but not to TVIP-H performance, $r(105) = .00$, N.S. Multiple regression techniques were then employed to test whether the relation between OL and PPVT-R performance was similar in both groups. A product vector representing the interaction term was created by multiplying a coded home language experience vector indicating OSH or ESH group membership by OL. Coefficients for the regression of PPVT-R and TVIP-H performance on OL in the two groups were not found to be significantly different from each other, $F(3, 101) = .85$, N.S., and $F(3, 101) = .57$, N.S., respectively. Analysis of variance (ANOVA) and analysis of covariance (ANCOVA) in the case of PPVT-R performance (OL with four levels treated as a covariate) were then performed to determine any significant effects of home language experience on the children's receptive vocabulary performance.

PPVT-R.—PPVT-R performance of both OSH and ESH group children was found to be significantly below that of the norming sample (z's = 7.89 and 3.24, $p < .005$, respectively). Even the ESH children, who as a group were of higher SES than Dunn's sample, scored below the mean. At the same time, ESH group children outperformed the OSH group children on the PPVT-R. The original means for the two groups were found to reflect differences on the covariate OL, $F(1, 101) = 21.21$, $p < .001$. However, as can be seen in Table 2, even after the means were adjusted to control for differences on OL, the ESH group mean was still significantly higher than the OSH group mean. The PPVT-R score for the ESH group remained within 1 SD of the mean for the standardization sample, while the score for the OSH group reached 2 SD below the same mean.

TVIP-H.—In contrast to the results on the PPVT-R, the TVIP-H performance of both OSH and ESH group children was comparable to that of the Spanish-speaking norming sample (z's = 1.26 and 1.40, respectively, N.S.). In addition, OSH and ESH group TVIP-H scores reflect comparable performance, $F(1,104) = .02$, N.S. While ESH group children were of higher occupational status than were TVIP-H norming sample children, OSH group children were roughly comparable to the Spanish Peabody norming sample. In any event, there was no effect of OL on the TVIP-H performance of our samples.

Since both tests were given at the same sitting, it was of interest to confirm whether scores on the second test were being inflated by the children remembering answers from the first test on those items that were comparable. Such an effect would be most pronounced, we reasoned, for cognates, which accounted for about 25% of the translation equivalents. These TEs, owing to their common roots, have similar pronunciation and meaning in both languages and hence may be easier to guess correctly if only one of the words is known to the child. Analysis of the percent right in both languages for approximately 15% of the sample revealed that these cognates were not "automatic doublets" for the children. The cognates were right in both languages about 68% of the time, compared to 67% for all words. Since the effect was so small on test items where

it might be expected to have the most effect, it is unlikely that the phonological similarity of cognates played any role in test performance.

Translation Equivalent (TE) Analysis

An item analysis of the TEs that OSH and ESH group children encountered shows that these groups of children knew TE words in both languages 59% and 67% of the time, respectively; knew words in neither language 20% and 14% of the time; knew only the English word 10% and 12% of the time; and knew only the Spanish word 12% and 8% of the time (see Table 3).

TE performance was examined to determine whether there was an effect of home language experience on the TE results. Given the differences found between OSH and ESH group children on OL, the possible effect of this variable on TE performance was examined first.

Pearson product-moment correlation coefficients were run to determine the relations between OL and percent right in English and Spanish (PRES), percent wrong in English and Spanish (PWES), percent right in English Only (PREO), and percent right in Spanish Only (PRSO) performance. OL was found to be related to PRSO, $r(95) = -.18$, $p = .01$, but not to PRES, $r(95) = .04$, N.S., PWES, $r(95) = -.03$, N.S., or PREO, $r(95) = .14$, N.S. Using multiple regression techniques, the relation between OL and PRSO was found to be similar in both language groups, $F(3, 90) = 1.29$, N.S. Analysis of variance (TE × language group), and, in the case of PRSO, analysis of covariance (TE × language group with four levels of OL treated as covariate), were then performed to determine any significant effects of home language experience on the children's TE performance. As shown in Table 3, ESH group children had significantly more TE words right in both English and Spanish, and OSH group children had significantly more TE words wrong in both English and Spanish. However, OSH and ESH group children were found to be performing similarly with regard to singlet vocabulary.

It was deemed important to verify that the proportions of items that were answered correctly in Spanish only (PRSO) and in English only (PREO) were significantly larger than would have been expected by chance (i.e., by subjects with no singlet vocabulary). The probability of correct guessing on two translation equivalent, four-choice items is 1/4 × 1/4 = 1/16. The probability of getting an item right in Spanish but wrong in English is 3/16, and right in English but wrong in Spanish, an additional 3/16. The probability of missing both by guessing is 9/16.

When the TE data for the two groups were combined, the following values were obtained for each variable: words known in neither language (PWES) 16%, English singlets (PREO) 11%, Spanish singlets (PRSO) 8%, and "doublets" (PRES) 65%. The percent right in English and Spanish (PRES) is contrary to the chance expectation, indicating, not surprisingly, that the children have vocabulary knowledge in the two languages.

The chance model (1:3:3:9 PRES:PRSO:PREO:PWES) was then used to predict the relative number of items that would by chance be wrong in one language or wrong in both; that is, children with no true singlet knowledge would be expected by chance to get 3/16 of the remaining items (those that were not among the doublets) right in one language only, 3/16 right in the other language only, and 9/16 wrong in both.

TABLE 3

TRANSLATION EQUIVALENT PERFORMANCE FOR ONLY SPANISH AT HOME (OSH) AND ENGLISH AND SPANISH AT HOME (ESH) CHILDREN

	OSH		ESH			
	Mean	SD	Mean	SD	*F*	*df*
PRES	59	19	67	15	4.38*	1,94
PWES	20	13	14	11	4.17*	1,94
PREO	10	13	12	12	< 1	1,94
PRSO	11	15	7	9	1.14	1,93

NOTE.—These data are for the 91% of the children who had six or more TE presentations in the course of their PPVT-R/TVIP-H testing, 25 OSH group children and 70 ESH group children. Acronyms refer to percent right in English and Spanish, percent wrong in English and Spanish, percent right in English only, and percent right in Spanish only.

* $p < .05$.

According to this reasoning, the expected ratio of PRSO or PREO to PWES is 6:9 (or 2:3).

Since the observed ratio of PRSO or PREO to PWES is known (in fact, 19% to 16%, or nearly 5:4), it is possible with a z test of proportions to verify that the real subjects do *not* conform to the model represented by chance performance (2:3) of subjects without singlet knowledge. Far too many items were answered correctly in just one language, when compared with the number answered incorrectly in both (z = 2.54, p = .0113), for the results to represent chance performance. The conclusion is that the subjects undeniably had singlet vocabulary. The singlet performance thus suggests that at least a portion of bilingual children's lexical knowledge is distributed disjunctively between two languages.

It is important to note that the procedure used to demonstrate the existence of singlet vocabulary did not allow the bilingual children in our study to exhibit the full range of such items in their repertories. Standard administration of the PPVT-R and the TVIP-H called for the establishment of basals and ceilings, a translation equivalent pair being counted only if the child encountered a translation equivalent on both tests. Had children been shown all the same plates in one language as in the other, they would have undoubtedly had more singlets. The demonstration of the existence of singlets under these adverse conditions points to a larger phenomenon that needs to be taken into account.

Once again, the effect of cognates, words of similar pronunciation and meaning, among the TE opportunities was considered. In the data available to evaluate this issue, singlets occurred less often with cognates than with noncognates, 17% versus 23%. This suggests that the presence of cognates tends to decrease the occurrence of singlets and would, therefore, if anything, make the existence of singlets as a nonrandom phenomenon harder to demonstrate.

Discussion

Our evaluation of *both* the English and the Spanish vocabulary knowledge of Hispanic children contradicts the prevailing view of bilingualism as a risk factor in vocabulary development. Consistent with the findings reported for bilingual children by other researchers utilizing the PPVT (Argulewicz & Abel, 1984; Ben-Zeev, 1977a, 1977b; Rosenblum & Pinker, 1983; Serapiglia, 1978), the English vocabulary performance of children in the ESH group, those exposed to English and Spanish in the home, was found to be in the low-average range. However, the Spanish vocabulary performance of these bilingual children was comparable to that of the monolingual Spanish-speaking TVIP-H norming sample, suggesting the importance of evaluating bilingual children in *both* their languages. The adequacy of the Spanish receptive vocabulary was shown in both of our groups, one roughly comparable in occupational level and one higher in OL than the norming sample for the test. Further, OL was not related to the Spanish vocabulary scores of our samples, indicating that the expected advantage conferred by OL was not a major factor in these results.

Second, home language experience was found to be an important variable in bilingual vocabulary acquisition. Even though OSH and ESH group children were found to be functioning at the same level on the TVIP-H, the English vocabulary performance of ESH group children was significantly better than that of OSH group children. Thus, while the bilingual experience of ESH children did them little if any harm in Spanish when compared to their peers receiving only Spanish at home, the more balanced language input experience of ESH group children had the benefit of laying the groundwork for superior performance in English. These findings also serve as a reminder that investigators should take specific account of the degree of exposure to the two languages when assessing bilingual knowledge.

The average performance of ESH group children in both their languages may be related to the sociocultural context in which bilingual experience occurs in Miami. Spanish is the first language of 49% of the population; powerful Spanish-language media compete favorably with English newspapers, radio, and TV. People of Hispanic background participate at high levels of government, banking, and industry, and all socioeconomic levels are represented in the Hispanic sector. Thus, the bilinguality experienced by Miami Hispanics may be for the most part "additive"—that is, bilinguality based on an appreciation of both cultures and languages. It is under these circumstances that Lambert (1977) proposes that the learning of a majority language does not encroach on the heritage language.

One might still question whether the

children in our study, especially ESH group children, were not sacrificing some proficiency in Spanish as a result of learning English. Does the three-point deficit, while not statistically significant, signify a trend toward lower language performance in bilingual children? Careful consideration of the psychometric properties and assumptions of the test suggests that the Spanish vocabulary knowledge of both of our groups may in fact be greater than their TVIP-H scores indicate.

Tomayo (1987) has demonstrated that a test translation that preserves the meaning of the original items but not their relative frequency has poor concurrent validity with the original instrument. The same word may exist in two dialects but occupy a different frequency position in each. Fernández, Pearson, Umbel, Oller, and Molinet-Molina (1992) found evidence of just such a confound while testing Hispanic children on the TVIP-H in Miami. The order of difficulty of the TVIP-H items was found to be substantially different for Miami Hispanics from that derived from the norming sample (in Mexico and Puerto Rico) by Dunn. The order of difficulty of items is critical to performance on the TVIP-H. A ceiling is established when the child misses six of the last eight items tested. Thus, both OSH and ESH group children were at a disadvantage on the TVIP-H to the extent that harder words preceded easier ones because of a lack of equivalence across dialects.

Our discussion thus far has focused on the use of the PPVT-R and the TVIP-H as language achievement tests. As achievement tests, vocabulary measures estimate the knowledge of language-specific labels for referents. Based on the moderate to high correlations between vocabulary and fuller tests of language (Doyle et al., 1978; Hakuta, 1987), the vocabulary score becomes a more general estimate of skill in that language. However, vocabulary measures are also widely used as aptitude tests.

Vocabulary tests become *aptitude* tests on the strength of the well-documented link between vocabulary measures and intelligence tests. Indeed, vocabulary subsections of intelligence tests consistently show the highest correlations to the full-scale scores (Terman, 1918; Wechsler, 1974). As indicators of aptitude, vocabulary tests estimate the number of lexicalized concepts the child has—the number of concepts for which the child has a lexicalized representation. It follows, then, that any valid use of the bilingual child's vocabulary as an indicator of aptitude must include concepts known in both languages. Thus, inferences about aptitude that derive from the comparison to age norms will be compromised if the child knows some words in only one language or the other. Words lexicalized only in Language A will not be counted in the standard score for Language B, and vice versa.

Indeed, a statistically significant proportion of the translation equivalent vocabulary tested with the children proved to be words known in one language but not the other, suggesting that the children were in command of more vocabulary concepts than either the PPVT-R or TVIP-H scores reflect. Even children who were very weak in one of the languages tended to know some words in each language that they did not know in the other. Further research with children through the sixth grade (Umbel, 1991) indicates that the proportion of singlets increases as children grow older and as their language scores become more balanced. Our findings are consistent with those reported by Swain (1972), who showed that while the bilingual children in her study had a smaller range of options for expressing the negative in each language compared to monolingual development patterns, they actually had more options from the two languages combined.

These results answer the empirical question of whether singlets, words lexicalized in only one language, exist for this population. Using available instruments, we have been able to point out a specific way in which standardized tests do not fully assess the knowledge of bilingual children. To the extent that inferences about the bilingual child's cognitive functioning are to be based on lexical knowledge, one must look beyond performance in either language to a union of skills in the two languages. New instruments will have to be devised, however, in order to assess total lexical knowledge in bilingual children. Until such a measure is developed, one must interpret standardized vocabulary scores with caution, knowing that even two single-language scores give only part of the picture.

References

American Educational Research Association, American Psychological Association, American Council on Measurement (1985). *Standards for educational and psychological tests*. Washington, DC: American Psychological Association.

children in our study, especially ESH group children, were not sacrificing some proficiency in Spanish as a result of learning English. Does the three-point deficit, while not statistically significant, signify a trend toward lower language performance in bilingual children? Careful consideration of the psychometric properties and assumptions of the test suggests that the Spanish vocabulary knowledge of both of our groups may in fact be greater than their TVIP-H scores indicate.

Tomayo (1987) has demonstrated that a test translation that preserves the meaning of the original items but not their relative frequency has poor concurrent validity with the original instrument. The same word may exist in two dialects but occupy a different frequency position in each. Fernández, Pearson, Umbel, Oller, and Molinet-Molina (1992) found evidence of just such a confound while testing Hispanic children on the TVIP-H in Miami. The order of difficulty of the TVIP-H items was found to be substantially different for Miami Hispanics from that derived from the norming sample (in Mexico and Puerto Rico) by Dunn. The order of difficulty of items is critical to performance on the TVIP-H. A ceiling is established when the child misses six of the last eight items tested. Thus, both OSH and ESH group children were at a disadvantage on the TVIP-H to the extent that harder words preceded easier ones because of a lack of equivalence across dialects.

Our discussion thus far has focused on the use of the PPVT-R and the TVIP-H as language achievement tests. As achievement tests, vocabulary measures estimate the knowledge of language-specific labels for referents. Based on the moderate to high correlations between vocabulary and fuller tests of language (Doyle et al., 1978; Hakuta, 1987), the vocabulary score becomes a more general estimate of skill in that language. However, vocabulary measures are also widely used as aptitude tests.

Vocabulary tests become *aptitude* tests on the strength of the well-documented link between vocabulary measures and intelligence tests. Indeed, vocabulary subsections of intelligence tests consistently show the highest correlations to the full-scale scores (Terman, 1918; Wechsler, 1974). As indicators of aptitude, vocabulary tests estimate the number of lexicalized concepts the child has—the number of concepts for which the child has a lexicalized representation. It follows, then, that any valid use of the bilingual child's vocabulary as an indicator of aptitude must include concepts known in both languages. Thus, inferences about aptitude that derive from the comparison to age norms will be compromised if the child knows some words in only one language or the other. Words lexicalized only in Language A will not be counted in the standard score for Language B, and vice versa.

Indeed, a statistically significant proportion of the translation equivalent vocabulary tested with the children proved to be words known in one language but not the other, suggesting that the children were in command of more vocabulary concepts than either the PPVT-R or TVIP-H scores reflect. Even children who were very weak in one of the languages tended to know some words in each language that they did not know in the other. Further research with children through the sixth grade (Umbel, 1991) indicates that the proportion of singlets increases as children grow older and as their language scores become more balanced. Our findings are consistent with those reported by Swain (1972), who showed that while the bilingual children in her study had a smaller range of options for expressing the negative in each language compared to monolingual development patterns, they actually had more options from the two languages combined.

These results answer the empirical question of whether singlets, words lexicalized in only one language, exist for this population. Using available instruments, we have been able to point out a specific way in which standardized tests do not fully assess the knowledge of bilingual children. To the extent that inferences about the bilingual child's cognitive functioning are to be based on lexical knowledge, one must look beyond performance in either language to a union of skills in the two languages. New instruments will have to be devised, however, in order to assess total lexical knowledge in bilingual children. Until such a measure is developed, one must interpret standardized vocabulary scores with caution, knowing that even two single-language scores give only part of the picture.

References

American Educational Research Association, American Psychological Association, American Council on Measurement (1985). *Standards for educational and psychological tests*. Washington, DC: American Psychological Association.

Argulewicz, E. N., & Abel, R. R. (1984). Internal bias in the PPVT-R for Anglo-American and Mexican American children. *Journal of School Psychology*, 22, 299–303.

Beck, F., Black, F., & Doles, J. (1985). The concurrent validity of the Peabody Picture Vocabulary Test—Revised relative to the Comprehensive Tests of Basic Skills. *Educational and Psychological Measurement*, 45, 705–710.

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

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

Cummins, J. (1978). Cognitive development of children in immersion programs. *Canadian Modern Language Review*, 34, 855–883.

D'Amato, R., Gray, J., & Dean, R. (1987). Concurrent validity of the PPVT-R with the K-ABC for learning problem children. *Psychology in the Schools*, 24, 35–39.

Doyle, A., Champagne, M., & Segalowitz, N. (1978). Some issues in the assessment of linguistic consequences of early bilingualism. In M. Paradis (Ed.), *Aspects of bilingualism* (pp. 13–21). Columbia, SC: Hornbeam.

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

Dunn, L., & Dunn, L. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: American Guidance Service.

Dunn, L., Padilla, E., Lugo, D., & Dunn, L. (1986). *Test de Vocabulario en Imágenes Peabody—Adaptación Hispanoamericana* [Peabody Picture Vocabulary Test—Latin American adaptation]. Circle Pines, MN: American Guidance Service.

Fernández, M. C., Pearson, B. Z., Umbel, V. M., Oller, D. K., & Molinet-Molina, M. (1992). Bilingual receptive vocabulary in Hispanic preschool children. *Hispanic Journal of the Behavioral Sciences*, 14, 268–276.

Figueroa, R. A. (1990). Best practices in the assessment of bilingual children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 2, pp. 93–106). Washington DC: National Association of School Psychologists.

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

Hakuta, K. (1987). Degree of bilingualism and cognitive ability in mainland Puerto Rican children. *Child Development*, 58, 1372–1388.

Hamers, J. F., & Blanc, M. A. (1989). *Bilinguality and bilingualism*. Cambridge: Cambridge University Press.

Lambert, W. E. (1977). Effects of bilingualism on the individual. In P. A. Hornby (Ed.), *Bilingualism: Psychological, social and educational implications* (pp. 15–28). New York: Academic Press.

Lambert, W. E., & Tucker, G. R. (1972). *Bilingual education of children: The St. Lambert experiment*. Rowley, MA: Newbury House.

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology*, 34, 388–399.

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs*, 76, 1–23.

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

Sattler, J. M., & Altes, L. M. (1984). Performance of bilingual and monolingual Hispanic children on the PPVT-R and the McCarthy Perceptual Scale. *Psychology in the Schools*, 21, 313–316.

Serapiglia, T. (1978). Comparison of the syntax and vocabulary of bilingual Spanish, Indian, and monolingual Anglo-American children. *Working Papers in Bilingualism*, 16, (ERIC Document Reproduction Service No. ED 165 477).

Skutnabb-Kangas, T., & Toukomaa, P. (1976). *Teaching migrant children's mother tongue and learning the language of the host country in the context of the socio-cultural situation of the migrant family* (Tutkimuksia Research Reports). University of Tampere, Finland, Department of Sociology and Social Psychology.

Swain, M. (1972). *Bilingualism as a first language*. Unpublished dissertation, University of California, Irvine.

Terman, L. M. (1918). The vocabulary test as a measure of intelligence. *Journal of Educational Psychology*, 9, 452–466.

Tomayo, J. M. (1987). Frequency of use as a measure of word difficulty in bilingual vocabulary test construction and translation. *Educational and Psychological Measurement*, 47, 893–902.

Umbel, V. M. (1991, January). *Measuring the receptive vocabularies of bilingual first, third, and sixth graders*. Paper presented at the XII Symposium on Spanish and Portuguese Bilingualism, Miami.

Wechsler, D. (1974). *Manual of the Weschler Intelligence Scale for Children—Revised*. New York: Psychological Corp.

Wilen, D., & Sweeting, C. (1986). Assessment of limited English proficient Hispanic students. *School Psychology Review*, 15, 59–75.