

# Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 8, Number 4, May, 2002

ISSN=1531-7714

## Language ability assessment of Spanish-English bilinguals: Future directions

*Ellen Stubbe Kester* and *Elizabeth D. Peña*

The University of Texas at Austin

Children from non-English speaking backgrounds are often misdiagnosed with language impairment due to a number of reasons. One of the primary reasons is that currently, there are limited diagnostic tools available that are designed for children who are exposed to two languages (Valdés & Figueroa, 1994). Current practices for assessment of language in bilinguals frequently involve the use of tests that are translated from English to the target language and/or tests designed for and normed on monolinguals. These currently available tools are not well suited for a bilingual population because they do not take into account the unique aspects of bilingual language acquisition. While the focus of this paper is on language assessment of bilinguals for the purpose of differentiating language impairment from typical language development, the issues presented have implications for all fields that include language as part of the assessment process, including IQ, educational, and achievement testing.

The objectives of this paper are to a) summarize relevant research on bilingual language development and discuss the implications for bilingual language assessment, b) discuss limitations in current language ability testing practices for bilinguals, c) propose future directions for the development of assessment tools and practices with bilinguals.

## Research on Bilingual Language Development: Implications for Assessment

Generally, the testing practices used today for bilinguals operate under the assumption that there is no difference in the language development of monolinguals and bilinguals. However, research in the area of bilingual language development suggests that bilinguals have different patterns of language development than monolinguals of either language (Grosjean, 1989). Consistent with the *Competition Model* (CM), Hernandez, Bates and Avila (1994) proposed that bilinguals use an amalgamation of strategies used by monolingual speakers. A key component of the CM model is that there are competing cues in any given language that help map meaning to utterances. The informational value of cues is determined by the frequency with which this type of information is available during decision-making processes and the frequency with which this type of information leads to a correct conclusion when it is used. As applied to language development, children test the use of different cues before they establish which cues best yield interpretations that are consistent with their environment.

Cross-linguistic studies based on this model indicate that the cues used to process and produce language efficiently are not the same across languages (MacWhinney & Bates, 1989). Examples of cues include word order and subject-verb agreement. In English, word order is relatively strict as compared to Spanish. The following sentence, "The boy (subject) hit (verb) the ball (object)" has a different meaning from "The ball (subject) hit (verb) the boy (object), because of word order cues. An English speaker would identify the subject (boy or ball) as the one hitting due to its position in the sentence. Spanish speakers, on the other hand, rely less on word order cues. For example, *El niño* (subject) *comió* (verb) *los frijoles* (object) (The boy ate the beans) has the same meaning as *Comió* (verb) *los frijoles* (object) *el niño* (subject). In comparison to English, Spanish has a complex verb system in which the verb stem provides cues about the subject, tense, and mood of the sentence. However, English verb morphology provides fewer cues about the subject. For example, the verbs *comí*, *comiste*, *comió*, and *comieron* are all represented by the verb *ate* (I ate, you ate, he ate, they ate, respectively) in English. Thus, bilingual children must learn how cues work within and between their two languages, creating a unique system of cues drawn from two languages (i.e., an amalgamated system). When children are developing two languages they often apply cues from L1 to L2 and from L2 to L1. Thus, bilingual children follow a different developmental course of language development in each of their languages in comparison to monolingual children. Language tests for bilinguals should reflect these differences in development.

## Limitations of Current Language Testing Practices for Bilinguals

Two common practices in the language assessment of bilinguals are translations of tests and the use of tests designed for monolinguals of the child's native language and/or second language. However, evidence that different linguistic cues are prominent in different languages and that bilinguals likely use an amalgamated cue system, suggests that translated tests and tests normed on monolinguals are likely to yield invalid estimates of language ability in bilinguals.

## ***Problems with Test Translation***

When tests are translated from one language to another, they do not retain their psychometric properties. Of particular interest in the assessment of language is the developmental order in which target features of the language are learned. Translating a test from one language to another -- typically from English -- may mean that items are organized by order of English difficulty, rather than reflecting the developmental order of the target language. The translated Spanish version of the Preschool Language Scale-3 (Zimmerman, Steiner, & Pond, 1993) provides an illustration. Restrepo and Silverman (2001) found several item difficulty discrepancies between the original English and the translated Spanish version when tested with predominately Spanish-speaking preschoolers. For example, items related to prepositions, which were relatively easy for English speakers, were more difficult for Spanish speakers. On the other hand, the "function" items were easier for the Spanish speakers in comparison to the English speakers.

The notion of cue validity can be used to examine development of semantic representation. Figueroa (1989) noted that words may generally represent the same concept but have variations and different levels of difficulty across languages, possibly due to their prominence, information load, and/or frequency. An illustration of this is found in a study of vocabulary test translations (Tamayo, 1987). When test items were translated from English to Spanish they differed in frequency of occurrence in each language. Because the Spanish translations were of lower frequency within Spanish, test scores obtained from Spanish speakers were lower compared to scores obtained from the original English version. However, when the vocabulary items were matched for their frequency of occurrence in the original and target language and matched for meaning, test scores obtained from Spanish and English speakers were equivalent.

Similarly, the context in which words are learned influences category development. Across different languages, the same general category may have different prototypical members, and different words may be associated with each language for the same situation. These contextual variations make translated vocabulary tests particularly vulnerable to imbalance. In a category generation task with bilingual four to six year-olds, Peña, Bedore, and Zlatic-Giunta (in press) found that for animals, children's three most frequent English responses were "elephant," "lion," and "dog," while in Spanish they used "caballo" (horse), "elefante" (elephant), and "tigre" (tiger) in these orders. Clearly, the circumstances under which children learn language affect their representation of language.

In addition to vocabulary differences, grammatical structure also affects the validity of test translation practices. For example, nouns are marked by gender in Spanish but not English, resulting in different cue values for each language. An English test translated to Spanish will miss aspects of Spanish, such as gender marking, that are not present in the English language. Furthermore, in Spanish, subject information is frequently carried in the verb, resulting in more complex verbs and less salient pronouns as compared to English. In English language assessment, pronoun omission is a hallmark of language impairment, yet this would not be true for Spanish. Thus, translated language tests may target inappropriate features for the target language, resulting in inaccurate assessment of language ability.

## ***Problems comparing bilinguals and monolinguals***

Bilingual school children generally fall into the category of circumstantial bilinguals. That is, their circumstances (often a Spanish-speaking home environment and an English-speaking or bilingual school environment) require them to use two languages. These different environments typically require different language content. The home environment likely promotes discussions of common family activities, such as cooking or trips to the store, while more academic topics, such as colors, numbers, and shapes, are highlighted in the school environment. As such, bilingual children will develop different vocabulary content for each language. From a testing perspective, this can result in underestimation of concept knowledge when testing in only one language at a time, or even when testing in both languages.

For example, Sattler and Altes (1984) examined typically developing three to six year-old bilingual Latino children's scores on the Peabody Picture Vocabulary Test-Revised and the McCarthy Perceptual Performance Scale. They found that the PPVT-R, whether administered in English or Spanish, yielded scores far below those of the norms, while all of the children were estimated to have normal intelligence based on their McCarthy scores.

Further investigation of the research on vocabulary development in bilinguals provides evidence of their use of a unique bilingual profile, and is consistent with the notion of an amalgamated rather than a "two monolinguals in one" system. A number of studies in the area of vocabulary acquisition illustrate that in early development, bilinguals learn unique words across their two languages, rather than learning two words (one in each language) for each concept. Pearson, Fernández, and Oller (1992) found that young bilinguals (8-30 months) often produced words for different concepts in each language, with few concepts labeled in both languages. Similarly Peña, Bedore, and Zlatic (in press) found that in a category generation task, bilingual children (ages 4-6 years) produced more unique words across Spanish and English (referred to as a *conceptual score*), in comparison to doublet (overlapped) words.

When monolinguals and bilinguals are compared on measures of vocabulary, differences become more apparent. Pearson, Fernández, and Oller (1993) used the Spanish and English versions of the MacArthur Communicative Development Inventory (1989) to estimate bilingual toddler's vocabularies. They found that when compared to monolingual norms in either language, their scores were low. However, when they compared the total number of unique words they produced across the two languages, their scores were more comparable to the monolingual norms.

Another example of findings of differential performance between monolinguals and bilinguals is with the Test de Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana (TVIP-H; Dunn, Padilla, Lugo, & Dunn, 1986). This version of the Peabody Picture Vocabulary Test (PPVT; Dunn, 1959) was normed on monolingual Spanish speakers outside of the U.S. mainland and then tested with bilingual Hispanics on the U.S. mainland. Results were that the bilinguals' scores were lower than those of the monolinguals (Dunn, 1988). Over age, the differences between monolinguals and bilinguals increased and coincided with schooling in English. Similarly, Umbel, Pearson, Fernandez, and Oller (1992) used the Peabody Picture Vocabulary Test-Revised (Dunn & Dunn, 1981) in English and the complementary Spanish version, the Test de Vocabulario en Imágenes Peabody (TVIP-H), to compare the receptive vocabularies of bilingual children (ages 5 years 11 months to 8 years 6 months) who were exposed to both Spanish and English in the home. Findings were that children on average responded correctly to 67% of the items in their age range in both languages, but that another 8% to 12% were known only in one of their two languages. Administration of this test in only one language -- even the "dominant" language -- would have led to an underestimation of vocabulary knowledge.

Conceptual scoring (Pearson, Fernández, & Oller, 1993) has been proposed as a more meaningful measure of the bilingual's conceptual knowledge. The system, which entails counting the concepts demonstrated (either through constructed or selected responses) in both languages and correcting for concepts shared in the two languages, results in a more valid representation of a bilingual child's knowledge of concepts.

## Future Directions

Item difficulty values, item discrimination, reliability, and validity are affected when tests are translated. For example, item difficulty values are affected when "equivalent" lexical items differ in frequency of occurrence (Tamayo, 1980). Less-frequent words have higher difficulty, while more frequent words are generally easier. Similar patterns of changes in item difficulty are seen for items that address conceptual framework, grammatical structure, and specific social content. The documented differences in bilingual and monolingual language development provide evidence suggesting that use of translated tests or tests designed for monolinguals will result in questionable validity. Clearly, the psychometric properties of a test do not translate from one language to another, nor do they remain the same when the test is administered to a different audience than intended.

While improving translation practices and uses of tests designed for monolinguals is an important short-term goal, long-term goals should include the development of language tests designed for, and normed on, bilinguals. In order to achieve such a goal, future research is needed to better understand the development of semantic and syntactic language skills in bilinguals. We offer the following recommendations to test developers:

- **Sample domains broadly during the exploratory level of test development to ensure that concepts and linguistic features are appropriately represented for each language.** For example, tests of semantic language skills should explore a wide variety of semantic concepts, such as similarities and differences in objects, functions of objects, categorization, characteristic properties, word associations, and spatial relations. Tests of grammar should explore a wide variety of structures in both languages rather than focusing on only the structures the two languages have in common, or on only structures important in English. Clinically, these suggestions apply as well. Testing beyond the ceiling, using dynamic assessment, clinical interviewing, and feedback during or as a follow-up to assessment of bilinguals may help better estimate true language ability.
- **Use conceptual scoring systems to eliminate underestimation of ability.** When testing concepts, consider a bilingual child's conceptual system as a whole, rather than as two language-specific systems. Thus, a bilingual approach accounting for the commonalities and differences across two languages is recommended over two monolingual assessments. When different concepts are expressed across languages, all should be counted. An example of an attempt at considering two languages is the English/Spanish Bilingual Verbal Ability Tests (BVAT) (Cummins, Muñoz-Sandoval, Alvarado, & Ruef, 1998), which assumes that bilinguals have a unique linguistic configuration, rather than two language-specific configurations. The BVAT estimates a bilingual's verbal ability by measuring the linguistic knowledge common to the bilingual's two languages and the linguistic knowledge unique to each language.
- **Select an appropriate mix of item types to gain the maximal amount of information about language ability in each language.** Rather than try to balance item types across languages, consider that some types of items may be more appropriate targets in one language than the other. For example, an English grammar test might include more items related to pronouns than a Spanish test because pronouns are more salient in English, whereas a Spanish grammar test might include more items related to gender and number agreement.
- **When trying to balance concepts in different language versions of tests, consider the frequency of occurrence of the words.** There are a number of published materials on word frequency in different contexts available in both Spanish and English that can be used to ensure that "equivalent" terms are not only equivalent in meaning but in frequency (or difficulty) as well.

## References

Cummins, J., Muñoz-Sandoval, A.F., Alvarado, C.G., & M.L. Ruef (1998). *The Bilingual Verbal Ability Tests*. Itasca,

IL: Riverside.

Dunn, L. H. (1959). *Peabody Picture Vocabulary Test*. Circle Pines, MN: American Guidance Services.

Dunn, L. H. (1988). *Bilingual Hispanic Children on the U. S. Mainland: A Review of Research on Their Cognitive, Linguistic, and Scholastic Development*. Honolulu, HI: Dunn Educational Services.

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

Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986). *Test de Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana*. Circle Pines, MN: American Guidance Service.

Figueroa, R. (1989). Psychological testing of linguistic-minority students: Knowledge gaps and regulations. *Exceptional Children*, *56*, 145-148.

Grosjean, F. (1989). Neurolinguists, Beware! The Bilingual is Not Two Monolinguals in One Person. *Brain and Language*, *36*, 3-15.

Hernandez, A. E., Bates, E., & Avila, L. X., (1994). On-line sentence interpretation in Spanish-English bilinguals: What does it mean to be “in between”? *Applied Psycholinguistics*, *15*, 417-46.

*MacArthur Communicative Development Inventory*. (1989). San Diego: University of California, Center for Research in Language.

MacWhinney, B. & Bates, E. (Eds.). (1989). *The Crosslinguistic Study of Sentence Processing*. New York: Cambridge University Press.

Pearson, B. Z., Fernandez, M. C., & Oller, D. K., (1992). Measuring bilingual children’s receptive vocabularies. *Child Development*, *63*, 1012-1221.

Pearson, B. Z., Fernandez, M. C., & Oller, D. K., (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, *43*, 93-120.

Peña, E. D., Bedore, L. M., & Zlatic-Giunta, R. (in press). Development of categorization in young bilingual children. *Journal of Speech, Language, and Hearing Research*.

Restrepo, M. A., & Silverman, S. W. (2001). Validity of the Spanish Preschool Language Scale-3 for use with bilingual children. *American Journal of Speech-Language Pathology*, *10*, 382-393.

Sattler, J. M., & Altes, L. M. (1984). Performance of bilingual and monolingual Hispanic children on the Peabody Picture Vocabulary Test—Revised and the McCarthy Perceptual Performance Scale. *Psychology in the Schools*, *21*, 313-316.

Tamayo, J. (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., Pearson, B. Z., Fernández, M. C., & Oller, D. K. (1992). Measuring bilingual children’s receptive vocabularies. *Child Development*, *63*, 1012-1020.

Valdés, G., & Figueroa, R. A. (1994). *Bilingualism and Testing: A Special Case of Bias*. Norwood, NJ: Ablex.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1993). *Preschool Language Scale-3: Spanish Edition*. San Antonio, TX: Psychological.

Send editorial correspondence to:

Ellen Stubbe Kester  
Department of Communication Sciences and Disorders  
Jesse H. Jones Communication Center, CMA 7.214  
The University of Texas at Austin  
Austin, TX 78712-1089

e-mail: [stubbe.kester@mail.utexas.edu](mailto:stubbe.kester@mail.utexas.edu)

**Descriptors:** Bilingualism; \* Hispanic Americans; \* Student Evaluation; \* Second Languages; Bilingualism; \* Gifted; \* Hispanic Americans; \* Screening Tests; \* Second Languages

**Citation:** Kester, Ellen Stubbe & Elizabeth D. Peña (2002). Language ability assessment of spanish-english bilinguals: future directions. *Practical Assessment, Research & Evaluation*, 8(4). Available online: <http://PAREonline.net/getvn.asp?v=8&n=4>.