

THE BILINGUAL RESEARCH JOURNAL
Winter 1996, Vol. 20, No. 1, pp. 55-68

COMPARATIVE VALIDITY OF THREE ENGLISH ORAL LANGUAGE PROFICIENCY TESTS

Fredrick A. Schrank
Riverside Publishing Company

Todd V. Fletcher
University of Arizona

Criselda Guajardo Alvarado
University of Houston, Clear Lake

Abstract

The validity of three English oral language proficiency tests was examined in terms of Cummin's BICS/CALP distinction. The tests studied included the *Idea Oral Language Proficiency Test* (IPT-I; Ballard, Tighe, & Dalton, 1989), the *Language Assessment Scales* (LAS; DeAvila & Duncan, 1991), and the *Woodcock Language Proficiency Battery--Revised* (WLPB-R; Woodcock, 1991). An examination of test content and pattern of correlations between subscales is used to describe the way oral language proficiency has been conceptualized and operationalized for each of the three oral language proficiency tests. These three tests measure similar and dissimilar aspects of oral language proficiency. Some of the similar and dissimilar aspects support the BICS/CALP distinction. Implications for evaluating oral language proficiency tests are discussed.

Oral language proficiency tests are frequently used to determine if language minority school-aged children are able to meet the academic task demands of monolingual instruction in English. As a consequence, the tests provide results that are used to establish or deny eligibility for

instruction in a language other than English. In practice, language-minority children scoring above a defined criterion on English oral language tests are presumed to be sufficiently proficient to participate, and function successfully, in a program of English-only instruction. Those falling below a defined criterion are eligible for a program of bilingual or English-as-a-second-language (ESL) instruction. Because these tests are used to make high stakes, irreversible decisions about the program of instruction in which a child is best suited to learn, it is important to critically examine their validity. One way to address the issue of validity is to compare the tests to one another in terms of what each measures.

It has been suggested that language proficiency may have two aspects that can be differentiated by context. For example, language proficiency has been defined as "the student's ability to communicate in an informal social setting *as well as* the ability to function in a more formal, cognitively demanding academic setting" (Hamayan, Kwiat, & Perlman, 1985, p. 21, italics ours). Cummins (1984) is often credited with formalizing a distinction between these two aspects of language proficiency. He described one aspect of language proficiency as *basic interpersonal communication skills* (BICS), or a repertoire of communication skills developed and used in everyday social situations. He described the other aspect of language proficiency as *cognitive-academic language proficiency* (CALP), or a type of language facility developed through, and needed for success in, formal classroom settings. His distinction has been posited to explain why teachers report that some language-minority students appear to be able to communicate in English but lack proficiency in language-related academic learning.

This article contains a description of the content for each of the subtests that constitute the English forms of the *Idea Oral Language Proficiency Test* (IPT-I; Ballard, Tighe, & Dalton, 1980), the *Language Assessment Scales* (LAS; DeAvila & Duncan, 1991), and the *Woodcock Language Proficiency Battery-Revised* (WLPB-R; Woodcock, 1991). Each test's reported concurrent validity data are summarized. It presents the results of an oral language concurrent validation study among these tests. Similarities and differences in test content are shown to exist. Cummins' distinction between BICS and CALP is used to explain some of the differences.

Method

Subjects. Subjects in this study included 77 bilingual kindergarten pupils in the Tucson, Arizona area and 119 bilingual grade 2 students in the Houston and Pasadena, Texas area.

Instruments. Idea Oral Language Proficiency Test (IPT-I). The IPT-I assesses four basic areas of English oral language proficiency: vocabulary, comprehension, syntax, and verbal expression. The instrument contains a wide range of items such as the ability to state name and age, identify common clothing, animals, and food, and identify common modes of transportation and household items. They also measure the ability to use present tense verbs and conjunctions, understand comparative and quantitative concepts, ask past tense questions, and use the conditional tense of verbs.

The IPT-1 manual presents evidence of concurrent validity as established through a study of teacher rating of oral language proficiency and IPT-I results. The IPT-I yields a total score for designation of students into three categories: Non-English Speaking, Limited English Speaking, and Fluent English Speaking.

Language Assessment Scales (LAS). The LAS is a series of tests for assessment of reading, writing, listening, and speaking proficiency of students in grades K to 12 whose home language is other than English. The Pre-LAS assesses the oral language skills of kindergarten and first grade students. The LAS-O assesses the listening and speaking ability of students in grades 1 through 12. The LAS R/W assess the reading and writing skills of students in grades 2 through 12. (The reading and writing measures were not included as part of this study.)

The Pre-LAS and LAS-O assess oral language proficiency through a sampling of phonology (phonemes, stress, rhythm, and intonation), lexicon (the "words" of the language), syntax (the rules for comprehending and producing meaningful utterances) and pragmatics (the appropriate use of language to obtain specific goals).

The Pre-LAS contains six sub-tests:

- *Simon Says* is a receptive test of ability to follow instructions.
- *Choose a Picture* is a receptive test of the child's understanding of simple oral classroom instructions and language used to talk about relationships, likenesses, and differences.

- *What's in the House* is an expressive test of ability to provide labels for common household objects, including articles of clothing, eating utensils, and furniture.
- *Say What You Hear* assesses receptive and expressive language ability through the repetition of orally-presented stimulus sentences.
- *Finishing Stories* is a test of ability to supply an appropriate clause to complete a compound or complex sentence.
- *Let's Tell Stories* is a test of ability to listen to stories and retell them.

The LAS-O contains four subtests:

- *Vocabulary* is a measure of the ability to produce object labels and other words in context, including the ability to produce antonyms for adjectives or adverbs and the "ing" form of verbs commonly used in conversation.
- *Listening Comprehension* is a measure of the ability to understand everyday conversations.
- *Story Retelling* is a test of ability to listen to stories and retell them.
- *Pronunciation* measures auditory discrimination of minimal pair items and ability to listen to and repeat specific phonemes embedded in words, phrases, and short sentences.

The LAS manual presents evidence of concurrent validity as established through a study of teacher rating of oral language proficiency and LAS-O results, and moderate correlations between the LAS-O and the *Bilingual Syntax Measure* (Burt, Dulay, & Hernandez-Chavez, 1978) and the *Basic Inventory of Natural Language* (Herbert, 1979).

The LAS-O yields a total score for designation of students into three language proficiency categories: Fluent Speaker, Limited Speaker, and Non-Speaker.

Woodcock Language Proficiency Battery-Revised (WLPB-R). The WLPB-R is designed to provide an overview of a subject's language skills in English and Spanish, to diagnose language abilities, to identify students for English as a second language instruction, and to plan broad

instructional goals for developing language competencies. The instrument is appropriate for individuals aged 2 to over 90 years of age. For interpretive purposes, each WLPB-R provides cluster scores for Broad Ability, Oral Language Ability, Reading Ability, and Written Language Ability. (The broad ability, reading, and writing measures were not included as part of this study.)

The WLPB-R Oral Language Cluster contains five subtests:

- *Memory for Sentences* measures the ability to remember and repeat phrases and sentences presented auditorily.
- *Picture Vocabulary* measures the ability to name familiar and unfamiliar pictured objects.
- *Oral Vocabulary* measures knowledge of synonyms and antonyms.
- *Listening Comprehension* measures the ability to comprehend a passage and supply the single word missing at the end in an oral cloze procedure. The test begins with simple verbal analogies and associations and progresses to a higher level of comprehension involving the ability to discern implications.
- *Verbal Analogies* measures the ability to comprehend and verbally complete a logical word relationship. Although the vocabulary remains relatively simple, the relationships among the words become increasingly complex.

The WLPB-R manual presents evidence of concurrent validity as established through moderate to strong correlations between the WLPB-R and measures of cognitive abilities, achievement, and language proficiency, including the *Wechsler Intelligence Scale for Children-Revised* (Wechsler, 1974), the *Stanford-Binet Intelligence Scale-Fourth Edition* (Thorndike, Hagen, & Sattler, 1986), the *Kaufman Assessment Battery for Children* (Kaufman & Kaufman, 1983), the *McCarthy Scales of Children's Abilities* (McCarthy, 1972), the *Kaufman Tests of Educational Achievement* (Kaufman & Kaufman, 1985), the *Peabody Individual Achievement Test* (Dunn & Markwardt, 1970), the *Test of English as a Foreign Language* (Educational Testing Service, 1987), the *Clinical Evaluation of Language Fundamentals-Revised* (Semel, Wug, & Secord, 1987), the *IDEA Oral Language Proficiency Test I* (Ballard, Tighe, & Dalton, 1980) and the *Test of Written Language* (Hammill &

Larsen, 1983). The WLPB-R Oral Language yields a cluster score for determination of oral language proficiency.

The Language Rating Scale. The LRS (see Figure 1) is a teacher rating of language proficiency. The LRS was obtained for use from the Houston Independent School District. It consists of a Likert-type rating for six areas of language ability, including sentence structure, vocabulary ability, word recall, telling stories, idea formulation, and speech production. For the purposes of this study, it is included as evidence of cognitive-academic language proficiency as measured by teacher judgment. (Item F, Speech, was not included in the data set, as this was presumed to be a measure of speech fluency).

Figure 1

**Houston Independent School District
Language Rating Scale**

English
Spanish

NAME _____ DOB _____ CAMPUS _____ IDF# _____
DATE _____

1	2	3	4	5
A. SENTENCE STRUCTURE: Always uses incomplete sentences with grammatical errors	Frequently uses incomplete sentences and/or numerous grammatical errors	Uses correct grammar; few errors of omission or incorrect use of prepositions, verb tenses or pronouns	Above average oral language; rarely makes grammatical errors	Always speaks in grammatically correct sentences
B. VOCABULARY ABILITY: Always uses immature or improper vocabulary	Limited vocabulary including primarily simple nouns; few precise, descriptive words	Adequate vocabulary for age and grade	Above average vocabulary; uses numerous precise descriptive words	High level vocabulary; always uses precise words to convey message; uses abstractions

C. RECALLING WORDS: Unable to call forth the exact words	Often gropes for words to express himself	Occasionally searches for correct word but adequate for age and grade	Above average ability; rarely hesitates on the word	Always speaks well; never hesitates or substitutes words
D. TELLING STORIES: Unable to relate isolated facts	Has difficulty relating ideas in logical sequence	Average ability to tell stories	Above average; uses logical sequence	Exceptional ability to relate ideas in a logical and meaningful manner
E. IDEA FORMATION: Unable to relate isolated facts	Has difficulty relating ideas in logical sequence	Usually relates facts into meaningful ideas; adequate for age and grade	Relates facts and idea well	Outstanding ability in relating facts appropriately
F. SPEECH: Speech is always unintelligible and/or dysfluent	Speech is difficult to understand; sometimes dysfluent; draws attention to itself	Speech is intelligible; fluent; adequate for age	Speech is usually clear and easy to understand	Speech is always distinct and fluent

Procedure

Kindergarten sample. Data were collected during the Fall of 1991 (September, October, and November) on 77 kindergarten students in five different classrooms of one elementary school in a large urban district. The subjects selected were identified as having a primary home language other than English on school enrollment forms and home language surveys. They were assessed for English and Spanish language proficiency using the PRE-LAS and the WLPB-R (the Spanish-language correlations are not included in this analysis). Three trained personnel administered the oral language proficiency tests in a counterbalanced order as a control for a possible ordering and learning effect. After the testing was completed, the Language Rating Scale was explained to the kindergarten teachers and they were asked to complete a form for each student in the study.

Grade 2 sample. Data for the Grade 2 sample were collected during the Fall of 1991 on 120 Grade 2 students in two large urban school districts from 27 elementary schools. The subjects were identified as having Spanish spoken at home on school enrollment forms and on home language surveys. The students were assessed on both English and Spanish forms of the WLPB-R, LAS-O (Form C), and the IPT-I (the Spanish correlations were not included in this analysis). Five trained personnel administered the oral language proficiency tests in a counterbalanced order as a control for possible ordering and learning effect. After the testing was completed, the LRS was explained to the teachers and they were asked to complete a form for each of their students. Bilingual (English and Spanish) teachers were asked to score their students' language proficiency in English and Spanish.

Results and Discussion

Table I presents the total test correlations for the kindergarten sample. At the Kindergarten level, the WLPB-R and the Pre-LAS show strong concurrent validity (.91). The WLPB-R shows the stronger correlation with the teacher's rating of language proficiency (LRS).

Table 1
Kindergarten Test Correlations

	WLPB-R Oral lang	PRE-LAS Total	Language Rating Scale
WLPB-R Oral lang.	1.00		
(n)	(74)		
PRE-LAS Total	.91	1.00	
(n)	(70)	(73)	
LRS Total	.80	.74	1.00
(n)	(62)	(65)	(65)

Table 2 presents the total test correlations for the Grade 2 sample. At the Grade 2 level, the WLPB-R, LAS, and IPT-I show evidence of concurrent validity (each .86), and the WLPB-R shows the strongest correlation with the teacher's ratings of language proficiency.

Table 2
Grade 2 Test Correlations

	WLPB-R Oral lang.	LAS Total	IDEA Total	Language Rating Scale
WLPB-R Oral lang.	1.00			
(<i>n</i>)	(119)			
LAS Total	.86	1.00		
(<i>n</i>)	(113)	(113)		
IDEA Total	.86	.75	100	
(<i>n</i>)	(119)	(114)	(119)	
LRS Total	.80	.76	.68	1.00
(<i>n</i>)	(107)	(102)	(107)	(107)

Correlations among Pre-LAS or LAS and WLPB-R subtests were derived to determine if the subtests from these two batteries were measuring similar or dissimilar aspects of oral language proficiency. The IPT-I does not provide subtest information. The LRS was not analyzed for this purpose.

Table 3 (Appendix) presents the English subtest correlations for the kindergarten sample. Correlations among the WLPB-R Oral Language subtests range from .56 to .86 with a median correlation of .72 (based on 10 intercorrelations). Correlations among PRE-LAS subtests range from .55 to .93 with a median correlation of .75 (based on 15 intercorrelations).

A high correlation between a subtest of one battery and a subtest of another battery provides evidence that those subtests measure a similar ability (or results from a common underlying factor). Among the correlations between the WLPB-R Oral Language and PRE-LAS subtests, the highest correlations include .90 between the WLPB-R Picture Vocabulary and the PRE-LAS Test 3: What's in the House? and .90 between the WLPB-R Listening Comprehension and PRE-LAS Test 3: What's in the House? scores. Low intercorrelations among subtests within a test provide evidence that the subtests measure different aspects of a broad ability. The lowest correlation is .36, between the WLPB-R Oral Vocabulary and the PRE-LAS Simon Says.

Table 4 (Appendix) presents the English subtest correlations for the Grade 2 sample. Correlations among the WLPB-R Oral Language subtests range from .76 to .90 with a median correlation of .86 (based on

10 intercorrelations). Correlations among LA S-0 subtests range from .39 to .95 with a median correlation of .81 (based on 6 intercorrelations). This result is due to the rather low correlations of the LAS-O Pronunciation subtest with the other LAS-O subtests.

Among the correlations between the WLPB-R Oral Language and LAS-O subtests, the highest correlations include .88 between the WLPB-R Listening Comprehension and the LAS-O Vocabulary and .86 between the WLPB-R Picture Vocabulary and LAS-O Vocabulary scores. The lowest correlation is .44, between the WLPB-R Verbal Analogies and the LAS-O Pronunciation.

The test correlations obtained from this study provide overall evidence of concurrent validity among the three tests. The IPT-I, LAS-0, and WLPB-R Oral Language total scores correlated highly at the Kindergarten level and well at the Grade 2 levels.

Differences among correlational patterns were found, however, when the subtests from one test were compared to those of another. For example, at the Kindergarten level, these correlations provide evidence to suggest that the WLPB-R Oral Vocabulary and the Pre-LAS Simon Says subtests measure different aspects of oral language proficiency (correlation .36). The LA 5-0 Simon Says is a receptive test of ability to follow instructions; this may be explained as a component of BICS. The WLPB-R Oral Vocabulary subtest measures knowledge of synonyms and antonym 5; this may be explained as a component of CALP. At the Grade 2 level, differences were also found. The WLPB-R Verbal Analogies and LAS-O Pronunciation appear to be measuring different aspects of language proficiency, as suggested by a low correlation (.44). The LAS-O Pronunciation test measures auditory discrimination of minimal pair items and ability to listen to and repeat specific phonemes embedded in words, phrases, and short sentences. This is a component of communicative language, and can be interpreted as a component of BICS. The WLPB-R Verbal Analogies measures the ability to comprehend and verbally complete a logical word relationship. This is an ability necessary for success in school, and can be described as a component of CALP.

The Language Rating Scale (LRS) was used as a teacher rating of each student's language proficiency. At both the Kindergarten and Grade 2 levels, the WLPB-R Oral Language correlated most strongly

with the teacher's ratings of language proficiency. This may also provide evidence that the WLPB-R Oral Language tests measure CALP better because these tests showed the strongest relationship with the teacher's ratings of English language proficiency.

Implications

In the final analysis, test validity can only be established within the context of the purpose of assessing oral language proficiency - to determine if children are able to meet the academic task demands of a mainstream English instruction classroom. Individuals responsible for selecting oral language proficiency tests for use in making high stakes, irreversible decisions regarding service eligibility either for instruction in a language other than English or a transition to English-only instruction should carefully examine test content to ascertain if the test is measuring cognitively-demanding academic content, or CALP. Measures of BICS may lead educators to assume that a student possesses CALP. This is not an innocuous assumption, as it may lead to incorrect placement decisions, denial of services, or a failure to succeed with language-related academic learning tasks. If oral language proficiency tests are to be used to determine if language-minority school-aged children are able to meet the academic task demands of monolingual instruction in English, then the test's task demands should also be academic in nature. Test users should examine each test manual for evidence of concurrent validity to measures of cognitive ability and academic achievement, as well as to other measures of language proficiency. If such evidence is not presented, users have no way of knowing if the test correlates well with established measures of cognitive-academic language proficiency.

References

- Ballard, W. S., Tighe, P. L., & Dalton, E. F. (1980). *IDEA Oral Language Proficiency Test I*. Brea, CA: Ballard & Tighe.
- Burt, M. K. Dulay, H. C., & Hernandez Chavez, E. (1978). *Bilingual Syntax Measure*. San Antonio, TX: The Psychological Corp.
- Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Austin, TX: Pro-Ed.

- De Avila, E. A., & Duncan, S. E. (1991). *Language assessment scales*. San Rafael, CA: Linguametrics.
- Dunn, L. IA., & Markwardt, F. C. (1970). *Peabody Individual Achievement Test*. Circle Pines, MN: American Guidance Service.
- Educational Testing Service. (1987). *Test of English as a foreign language*. Princeton, NJ: Author.
- Hamayan, B., Kwiat, J., & Perlman, R. (1985). *The identification and assessment of language-minority students: A handbook for educators*. Illinois Resource Center, Arlington Heights, Illinois.
- Hammill, D. D., & Larsen, S. C. (1983). *Test of Written Language*. Austin, TX: Pro-Ed.
- Herbert, C. H. (1979). *Basic Inventory of Natural Language*. San Bernadino, CA.: CHECpoint Systems.
- Houston Independent School District. *Language Rating Scale*. Houston, TX
- Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S., & Kaufman, N. L. (1985). *Kaufman Tests of Educational Achievement*. Circle Pines, MN: American Guidance Service.
- McCarthy, D. A. (1972). *McCarthy Scales of Children's Abilities*. San Antonio, TX: Psychological Corporation.
- Semel, E., Wug, E., & Secord, W. (1987). *Clinical Evaluation of Language Fundamentals-Revised*. San Antonio, TX: Psychological Corporation
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale-Fourth Edition*. Chicago: Riverside.
- Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corporation.
- Woodcock, R. W. (1991). *Woodcock Language Proficiency Battery Revised*. Chicago: Riverside.

Table 3
Subtest Correlations for Kindergarten Sample

	Memory for Sentences	Picture Vocab- ulary	Oral Vocab- ulary	Listening Compre- hension	Verbal Analogies	Test 1: Simon Says	Test 2: Choose a Picture	Test 3: What's in a House	Test 4: Say What You Hear	Test 5: Finishing Stories	Test 6: Let's Tell Stories
WLPB-R											
Memory for Sentences (n)	1.00 (77)										
Picture Vocabulary (n)	.77 (77)	1.00 (77)									
Oral Vocabulary (n)	.56 (77)	.65 (77)	1.00 (77)								
Listening Comprehension (n)	.71 (77)	.86 (77)	.73 (77)	1.00 (77)							
Verbal Analogies (n)	.67 (77)	.77 (74)	.68 (74)	.83 (74)	1.00 (77)						
PRE-LAS											
Test 1: Simon Says (n)	.47 (73)	.64 (73)	.36 (73)	.53 (73)	.52 (70)	1.00 (73)					
Test 2: Choose a Picture (n)	.49 (70)	.70 (70)	.50 (70)	.69 (70)	.70 (67)	.83 (70)	1.00 (70)				
Test 3: What's in a House (n)	.69 (68)	.90 (68)	.62 (68)	.90 (68)	.77 (65)	.55 (68)	.74 (68)	1.00 (68)			
Test 4: Say What You Hear (n)	.71 (64)	.76 (64)	.53 (64)	.74 (64)	.71 (61)	.68 (64)	.75 (64)	.77 (64)	1.00 (64)		
Test 5: Finishing Stories (n)	.68 (73)	.85 (73)	.65 (73)	.87 (73)	.82 (70)	.63 (73)	.76 (70)	.88 (68)	.77 (64)	1.00 (73)	
Test 6: Let's Tell Stories (n)	.65 (73)	.86 (73)	.62 (73)	.87 (73)	.85 (70)	.60 (73)	.72 (70)	.85 (68)	.74 (64)	.93 (73)	1.00 (73)

Table 4
Subtest Correlations for Grade 2 Sample

	WLPB-R					LAS			
	Memory for Sentences	Picture Vocabulary	Oral Vocabulary	Listening Comprehension	Verbal Analogies	Vocabulary	Listening Comprehension	Story Retelling	Pronunciation
WLPB-R									
Memory for Sentences (n)	1.00 (119)								
Picture Vocabulary (n)	.82 (119)	1.00 (119)							
Oral Vocabulary (n)	.78 (119)	.88 (119)	1.00 (119)						
Listening Comprehension (n)	.78 (119)	.90 (119)	.90 (119)	1.00 (119)					
Verbal Analogies (n)	.76 (119)	.84 (119)	.90 (119)	.90 (119)	1.00 (119)				
LAS									
Vocabulary (n)	.72 (98)	.86 (98)	.78 (98)	.88 (98)	.80 (98)	1.00 (98)			
Listening Comprehension (n)	.65 (98)	.80 (98)	.74 (98)	.81 (98)	.76 (98)	.95 (98)	1.00 (98)		
Story Retelling (n)	.65 (98)	.81 (98)	.74 (98)	.82 (98)	.75 (98)	.94 (98)	.91 (98)	1.00 (98)	
Pronunciation (n)	.64 (41)	.53 (41)	.55 (41)	.72 (41)	.44 (41)	.71 (13)	.39 (13)	.46 (13)	1.00 (13)