

EAC West, New Mexico Highlands University, Albuquerque, March, 1996

Handbook of Spanish Language Proficiency Tests

Michael Guerrero, Ph.D.
and
Ann Del Vecchio, Ph.D.

Evaluation Assistance Center-Western Region
New Mexico Highlands University
Albuquerque, NM

March 1996

Table of Contents

- [Introduction](#)
 - [The US Spanish-speaking population at a glance](#)
 - [The legal mandate](#)
 - [Definitions](#)
 - [Language proficiency within an educational context](#)
 - [General nature of language proficiency tests](#)
 - [Challenges in the development of Spanish language proficiency tests](#)
 - [Summary](#)
 - [Organization of the Language Proficiency Test Handbook](#)
 - [Glossary](#)
 - [Organization of test information](#)
 - [Checklist](#)
 - [Basic Inventory of Natural Language \(BINL\)](#)
 - [Bilingual Syntax Measure I and II Spanish \(I & II\)](#)
 - [Spanish Idea Proficiency Tests \(IPT\)](#)
 - [Language Assessment Scales \(LAS\)](#)
 - [Woodcock Muñoz Language Survey](#)
 - [Summary and test descriptions matrix](#)
 - [References](#)
-

Introduction

The purpose of this handbook, like its predecessor (Del Vecchio & Guerrero, 1995), is to provide educators responsible for assessing the Spanish language proficiency of students with information about commercially available, standardized Spanish language proficiency tests. None of the tests are critiqued directly and no endorsement of any of the tests is intended. Our intent is to assist educators in making an informed Spanish language proficiency test adoption decision.

In addition to information about these tests, background information about Spanish language proficiency testing is included. Clearly, assessing the Spanish language proficiency of students is in many ways similar to assessing students' English language proficiency. On the other hand, there is a series of unique factors which pertain only to the assessment of Spanish language proficiency within the U.S. The *Introduction* section will address these two sets of issues.

The U.S. Spanish Speaking Population at a Glance

Any cursory review of research on the numbers of Spanish language origin children will quickly reveal the vitality and diversity of this sector of the U.S. population. Consider the following data:

- Drawing on the 1980 U.S. Census, Solé (1990) indicates that four out of every 10 Hispanic preschool children are either incipient bilinguals (i.e., understand some Spanish) or monolingual Spanish.
- Waggoner (1995), utilizing 1990 US Census data, indicates that there are an estimated 4.2 million home speakers of Spanish between the ages of 5 and 17. This represents approximately two-thirds of the total school age population who reportedly come from a household where a non-English language is used.
- Waggoner (1994) indicates that households reporting the use of Spanish with children between the ages of five and 17 can be found in every state in the U.S.
- The top ten states with the largest numbers of home speakers of Spanish in order from largest to smallest are: California, Texas, New York, Florida, Illinois, New Jersey, Arizona, New Mexico, Massachusetts, and Pennsylvania (Waggoner, 1995).
- Between 1981 and 1990, thousands of Spanish speaking immigrants came to the U.S. from various parts of the world including: Mexico, the Dominican Republic, Cuba, El Salvador, Argentina, Ecuador, Spain and other Caribbean, Central and South American countries (Figueroa and García, 1994).

These data reveal that the school age, Spanish language population in the U.S. is substantial, dispersed in different concentrations across the country, and linked to immigration from a number of different countries. In sum, the need to assess the Spanish language proficiency of school age children appears to be widespread. Nonetheless, the challenge of developing and using valid Spanish language proficiency measures, for reasons explained below, is a formidable one.

The Legal Mandate

Through the Office for Civil Rights, there is no explicit requirement for a local education agency (LEA) to assess the Spanish language proficiency of a student coming from such a language background. Only the English language proficiency (i.e., speaking, listening, reading and writing) of the learner coming from a non-English language background must be assessed (Roos, 1995). On the other hand, Roos does indicate that a local education agency would be well advised to offer a bilingual program whenever it is possible. Integral to offering a bilingual program is the need to assess the native language proficiency of the students.

A number of states and school districts have explicit procedures schools must follow to identify potential

LEP students. Part of these procedures may entail the assessment of the learner's Spanish language proficiency. Some states and school districts endorse the use of specific Spanish language proficiency tests. It is the responsibility of the reader to become informed about state or district policy with regard to the assessment of Spanish language proficiency.

The passage of the Improving America's Schools Act (IASA) will increase the need for educators to select and use measures of Spanish language proficiency, specifically under Title I and Title VII. Under *Section 1111, State Plans, (3) Assessments.*, Title I law indicates that yearly student assessments must be administered in at least math and reading or language arts. Further, Title I programs must provide for:

(iii) the inclusion of limited English proficient students who shall be assessed, to the extent practicable, in the language and form most likely to yield accurate and reliable information on what such students know and can do, to determine such students' mastery of skills in subjects other than English (emphasis added).

In addition, and under the same section, the newly legislated Title I guidelines state:

(5) LANGUAGE ASSESSMENTS.--Each State plan shall identify the languages other than English that are present in the participating student population and indicate the languages for which yearly student assessments are not available and are needed. . . .

This means that some Title I programs serving Spanish speaking students may have a need to at least assess the learners' math and reading or language arts achievement in the Spanish language. What is implied is that Title I programs will also need a Spanish language proficiency measure in order to first determine which students can best benefit from math and reading or language arts instruction and assessment activities in the Spanish language. To offer content area instruction and assessments in Spanish without having an indication of the learners' abilities to speak, understand, read, and write Spanish could render the use of Spanish content area assessments invalid.

Under the newly reauthorized IASA Title VII guidelines for developmental bilingual education programs, the evaluation requirements include that the grant recipient must report student gains in non-English language proficiency. In all likelihood, educators implementing dual-language bilingual education programs aimed at developing Spanish-English bilingualism will turn to one of the five standardized Spanish language proficiency tests described in this handbook to meet this evaluation requirement.

Finally, while some test users will be required to assess the Spanish language proficiency of their students due to some legal mandate, many local education agencies serving Spanish speaking students have no mandates. Nonetheless, there is an overwhelming consensus in the field of language minority education that educators must assess the native language skills of students in order to best serve these students (CCSSO, 1992). Clearly, a student who is literate in her native language will not require the same instructional program as a peer who is not literate. A language assessment must take place in order to make such a determination. The recent research by Collier (1995) underscores the interdependence among the student's native language development, acquisition of English, and academic success. Collier states:

To assure cognitive and academic success in a second language, a student's first language system, oral and written, must be developed to a high cognitive level at least through the elementary school years (1995, p. 3).

The implication of this statement is that the careful assessment of the student's native language proficiency must be carefully conducted over an extended period of time.

In sum, whether or not a local education agency must assess the Spanish language proficiency of a limited English proficient learner from such a home language background is contingent upon district, state or IASA policies. On the other hand, a local education agency serving Spanish speaking students with no binding requirement to assess the Spanish language proficiency of its students is urged to do so.

Definitions of Language Proficiency

It is a psychometric axiom that test developers must first define what it is they intend to measure. Unfortunately, it is at this point in the assessment of language proficiency that lack of a consensus begins. Language researchers have set forth definitions of language proficiency that vary dramatically. Consider the following brief but relatively recent definitions of language proficiency:

- Canale (1981) defines communicative competence (i.e., language proficiency) as consisting of four basic components: grammatical competence, sociolinguistic competence, discourse competence and strategic competence.
- Bachman (1990) defines language competence (i.e., language proficiency) as consisting of two general competencies; organizational competence which consists of grammatical and textual competence; and pragmatic competence which consists of illocutionary and sociolinguistic competence.
- Oller (1991) defines language proficiency as consisting of four semiotic capacities; the General Semiotic Capacity which regulates the lesser linguistic, kinesic and sensory-motor semiotic capacities.
- Canales (1994) defines language usage (i.e., language proficiency) as dynamic and contextually-based (varies depending upon the situation, status of the speakers, and the topic), discursive (requires connected speech), and requires the use of integrative skills to achieve communicative competence.

Valdés and Figueroa (1994) conclude that:

...what it means to know a language goes beyond simplistic views of good pronunciation, "correct" grammar, and even mastery of rules of politeness. Knowing a language and knowing how to use a language involves a mastery and control of a large number of interdependent components and elements that interact with one another and that are affected by the nature of the situation in which communication takes place. (p. 34)

Again, the complexity of language and the lack of a consensus as to the exact nature of language proficiency is critical for one fundamental reason. Each language proficiency test should be based on a defensible model or definition of language proficiency. The question becomes, on which definition? The consequence has been that language proficiency tests take on fairly distinct forms. Further, and while the test developer may indicate that a test is based on a particular model of language proficiency, it remains to be seen just how successfully the model was actually operationalized. In other words, describing the theoretical model of language proficiency in a technical manual does not mean that the test exemplifies the model.

Language Proficiency Within an Educational Context

While the debate among theoreticians continues concerning how to best define language proficiency, educators have set forth some fairly concrete descriptions of the kinds of language abilities (i.e., academic language proficiency) learners should have the opportunity to develop. Unfortunately, these language abilities have only been described for the English language, but they can obviously serve as benchmarks for the kinds of Spanish language skills educators should be moving learners toward. For example, the CCSSO (1992) defines English language proficiency in this way:

A fully English proficient student is able to use English to ask questions, to understand teachers, and reading materials, to test ideas, and to challenge what is being asked in the classroom. Four language skills contribute to proficiency as follows:

1. *Reading* - the ability to comprehend and interpret text at the age and grade-appropriate level.
2. *Listening* - the ability to understand the language of the teacher and instruction, comprehend and extract information, and follow the instructional discourse through which teachers provide information.
3. *Writing* - the ability to produce written text with content and format fulfilling classroom assignments at the age and grade-appropriate level.
4. *Speaking* - the ability to use oral language appropriately and effectively in learning activities (such as peer tutoring, collaborative learning activities, and question/answer sessions) within the classroom and in social interactions within the school. (p. 7)

Again, there is no legitimate reason why the academic Spanish language proficiency of learners should be construed in any different terms. Equally important, this definition should also help orient test developers who have assumed the responsibility of developing language proficiency tests for use in an educational setting. The implication is that test developers must design language measures that stem from the kinds of linguistic demands that underlie successful language use in an academic setting.

This is exactly what Valdés and Figueroa (1994) call for.

These researchers take the position that it is feasible to:

...identify the levels of demand made by such contexts and the types of language ability typical of native, monolingual English-speaking children who generally succeed in such contexts. From these observations, one could derive a set of criteria against which to measure the abilities of nonnative English-speaking children in order to decide whether to educate them in English or their home language. (p. 62)

Consequently, language proficiency tests which will be used to help make educational decisions should be intimately linked to the use of language within an educational context. Collier (1995) prefers to use academic achievement measures (i.e., standardized tests and performance assessments in language arts, reading, mathematics, science and social studies) used by schools to monitor learners' progress as measures of academic language proficiency. She refers to these measures as the "ultimate measures of academic proficiency in a second language" (p. 5).

Readers should be aware of the following. The greater the disparity is between the demands of language tasks on a language proficiency test and the language tasks of a typical classroom, the less valid and useful the test score will be. The propensity for making inappropriate educational decisions thus becomes greater.

General Nature of Language Proficiency Tests

Oller and Damico (1991) maintain that language proficiency tests can be associated with essentially three schools of thought. The first of these trends, the discrete point approach, was based on the assumption that language proficiency

...consisted of separable components of phonology, morphology, lexicon, syntax, and so on, each of which could be further divided into distinct inventories of elements (e.g., sounds, classes of sounds or phonemes, syllables, morphemes, words, idioms, phrase structures, etc.) (p. 82).

Oller and Damico describe language tests based on the discrete point approach in the following way:

Following the discrete point model, a test could not be valid if it mixed several skills or domains of structure (Lado, 1961). By this model, presumably the ideal assessment would involve the evaluation of each of the domains of structure and each of the skills of interest. Then, all the results could be combined to form a total picture of language proficiency. (p. 82).

A discrete point language proficiency test typically uses testing formats such as phoneme discrimination tasks where the test taker is required to determine whether or not two words presented aurally are the same or different (e.g., /ten/ versus /den/). A similar example might be a test designed to measure vocabulary which requires the test taker to point to a specified item in a picture or name an object.

The authors conclude that the weaknesses leading to the demise of such thinking centered upon evidence such as:

- the difficulty of limiting language testing to a single skill (e.g., writing) without involving another (e.g., reading);
- the difficulty of limiting language testing to a single linguistic domain (e.g., vocabulary) without involving other domains (e.g., phonology); and
- the difficulty of measuring language in the absence of any social context or link to human experience.

According to Damico and Oller (1991), these limitations gave rise to a second trend in language testing, the integrative or holistic approach. Basically, under this trend language proficiency could only be assessed "in a fairly rich context of discourse" (p. 83). This assumption was based on the belief that language processing or use entails the simultaneous engagement of more than one language component (e.g., vocabulary, grammar, gesture) and skill (e.g., listening, speaking). Following this logic, an integrative task might require the test-taker to listen to a story and then retell the story or to read the story and then write about the story.

The third language testing trend described by the Damico and Oller (1992) is referred to as pragmatic language testing. It differs from the integrative approach in one fundamental way; an ostensible effort is made to link the language testing situation with the test-taker's experience. As Oller and Damico (1991)

state, normal language use is connected to people, places, events and relations that implicate the whole continuum of experience and is always constrained by time or temporal factors. Consequently, pragmatic language tasks are intended to be as "real life" or authentic as possible.

In contrast to an integrative task, a pragmatic approach to language testing might require the test-taker to engage in a listening task only under the contextual and temporal conditions that generally characterize this activity. For example, if the test-taker is going to listen to a story and then retell the story, the following conditions might apply. From a pragmatic perspective, language learners do not generally listen to audio-taped stories; they more commonly listen to adults or competent readers read stories. In this sense a story-retell listening task which uses a tape-mediated story falls short of meeting pragmatic criteria. A pragmatic approach to story retelling might take on the following features:

- normal visual input is provided (e.g., the reader's gestures, the print on the page, an authentic number of story-linked pictures in the text);
- time is managed differently in that the learner may have opportunities to ask questions, make inferences, or react in a normal way toward the content of the story; and
- the story, its theme, the reader, and the purpose of the activity form part of the learner's experience.

Oller and Damico (1991) make an interesting observation regarding the power of pragmatic language testing. The researchers state that:

What was more important about pragmatic tests, and what is yet to be appreciated fully by theoreticians and practitioners (e.g., Spolsky, 1983), is that all of the goals of discrete point items (e.g., diagnosis, focus, isolation) are better achieved in the full rich context of one or more pragmatic tests . . . As a method of linguistic analysis, the discrete point approach had some validity, but as a practical method for assessing language abilities, it was misguided, counterproductive, and logically impossible. (p. 85)

In other words, if the intent is to measure the learner's proficiency in the areas of grammar, vocabulary, and pronunciation, for example, this is best achieved through a pragmatic language approach as opposed to a discrete point approach.

In sum, language proficiency testing approaches tend to fall into one or more of the three trends briefly outlined above. It is, however, the pragmatic language approach which seems most effectively to meet the demands of educators as set forth by federal education mandates, state education staff and academicians previously described. Nonetheless, educators will be limited to the use of currently available tests which may or may not measure language proficiency in a pragmatic manner within an educational context.

Challenges in the Development of Spanish Language Proficiency Tests

It is important for the reader to understand that the measurement of Spanish language proficiency presents additional, language-specific challenges which have not been directly dealt with up to this point. In general, the central difficulty with developing Spanish language proficiency tests concerns the norming of these tests. These challenges are presented below.

The variety of Spanish that each distinct group of immigrants bring with it presents a particular challenge

for standardized language measurement. For example, while a child from Spain is likely able to communicate with a peer from El Salvador, there will be differences between the two varieties of Spanish at virtually every linguistic juncture (e.g., phonological, syntactic, lexical, pragmatic). Consequently, the establishment of a Spanish language standard by which a learner's language ability will be measured becomes elusive. That is, it is difficult to develop a Spanish language test that will not place some students at a disadvantage because of the variety of Spanish they speak.

To address this challenge, some test developers have sought to norm their test on various Spanish speaking populations outside of the U.S. In this way, test items that appear to be biased toward different Spanish speaking populations can be identified, modified or deleted. While this process appears appropriate, it entails two major test development shortcomings. First, a child living in Mexico resides in a substantially different sociolinguistic milieu than the Spanish speaking child residing, for example, in Las Vegas, Nevada. The latter child does not have the same degree of access to the same social settings and speakers of the Spanish language as the child from Mexico (see Wong Fillmore, 1989 for a more detailed explanation of these factors).

In addition, the child residing in the US begins to have contact with the English language. This fact makes the child residing in the US potentially bilingual in Spanish and English and not Spanish monolingual like the learners from the norming sample. To compare a child's language proficiency who is in contact with two languages to a child's language proficiency who is in contact with only one language is questionable science (Valdés and Figueroa, 1994). The cognitive, linguistic, psychological, and social processes of becoming bilingual are not the same for a bilingual as they are for a monolingual. These two related factors generally influence the Spanish language development of the student residing in this country which makes the use of imported monolingual Spanish speaking norms of questionable validity.

Some test developers have sought to address this problem by norming their test on bilingual populations in the US. Again, this strategy seems appropriate but nonetheless includes some pitfalls. First, there is no consensus among theoreticians, researchers or educators concerning what it means to be "bilingual" (Skutnabb-Kangas, 1988). For example, is a child bilingual if she speaks English but only understands some spoken Spanish depending on the topic? What we do know about a bilingual person is that there are varying degrees of proficiency (e.g., native, near native, etc.) in either language and that each language is generally used in distinct domains (e.g., schooling, conducting business, engaging in religious activities).

The further problem with this approach is of a more practical nature. Where in the US are there stable bilingual communities to be found which use their Spanish language for educational purposes? Figueroa and García (1994) cite a recent study conducted by the US Department of Education in which the following relevant findings were reported:

- only 17% of schools provide a significant degree of primary language instruction;
- instructional aides provide a large amount of the instructional and translation services; and
- the majority of teachers serving Spanish-speaking pupils have no proficiency in Spanish.

Given these educational conditions, it is not surprising that sociolinguists are in general agreement that the use of Spanish in the U.S. is not stable across generations of speakers. Based on 1990 Census data, Hernández Chávez (1995) demonstrated that the use of Spanish among the school age Spanish language origin population decreased from 1980 to 1990. The problem for test developers becomes finding school age

youth who are not gradually shifting from Spanish to English over time.

While no one can deny that there are some bilingual communities in different areas of the U.S., and a few successful bilingual programs, test developers are still hard pressed to find adequate numbers of learners on whom to norm their tests that are given ample opportunity to develop their academic Spanish language proficiency. The implication is that norms established by using bilingual students in the U.S. run the risk of being impoverished compared to monolingual Spanish or English norms. However, this is due largely to the educational system's language policies and practices which generally discourage the use and continued development of non-English languages.

In sum, we have informed the readers of various factors which challenge test developers to set forth a valid Spanish language proficiency test for use in an educational setting. We encourage the reader to pay close attention to the manner in which each test developer approaches norming their particular test. In this way, educators serving Spanish speaking students can more fully understand the strengths and weaknesses of each test and make a more informed test adoption decision.

Summary

In the *Introduction*, it was clearly stated that the intent of this handbook is to provide descriptive information on five widely-used Spanish language proficiency tests. It was not to critique the five Spanish language proficiency tests described. The following information is intended to inform the test user (i.e., educators) about the limitations of these tests in general and to help explain these limitations.

First, it has already been stated that language proficiency tests need to be based on a particular theory or model of language proficiency. However, there is no consensus among researchers regarding the nature of language proficiency. The result has been the development of language proficiency tests which differ in fundamental ways from one another.

Second, and related to the design of language proficiency tests, there may be a propensity for test developers to use a discrete point approach to language testing. Valdés and Figueroa (1994) state:

As might be expected, instruments developed to assess the language proficiency of "bilingual" students borrowed directly from traditions of second and foreign language testing. Rather than integrative and pragmatic, these language assessment instruments tended to resemble discrete-point, paper-and-pencil tests administered orally. (p. 64)

Figueroa and García (1994) place the matter into proper perspective by stating:

Scores derived from one, two, or three hours of small responses to small stimuli (test items) do not account for much of the real-life functioning, nor for the situated and unpredictable nature of human behavior. (p. 18-19)

Consequently, and to the degree that the above two points are accurate, currently available language proficiency tests may not only yield questionable results about student's language abilities, but the results are based on an impoverished model of language testing.

Third, the results generated by Spanish language proficiency tests may be further suspect due to the complexities involved in norming the test either abroad or within the educational context of the U.S.

In closing this section of the handbook, consider the advice of Spolsky (1984):

Those involved with language tests, whether they are developing tests or using their results, have three responsibilities. The first is to avoid certainty: Anyone who claims to have a perfect test or to be prepared to make an important decision on the basis of a single test result is acting irresponsibly. The second is to avoid mysticism: Whenever we hide behind authority, technical jargon, statistics or cutely labeled new constructs, we are equally guilty. Thirdly, and this is fundamental, we must always make sure that tests, like dangerous drugs, are accurately labeled and used with considerable care. (p. 6)

[\(table of contents \)](#)

Organization of the Language Proficiency Test Handbook

This section of the handbook contains a brief description of the way the handbook is organized and our rationale for the inclusion of specific information about the tests. We chose specific kinds of information to include in order to assist districts, schools, and programs to make an informed test adoption decision. We have provided information about the test materials, cost, test administration and other pragmatic concerns. This information is provided as a "how to" guide. It is meant to be neither a critique nor an endorsement of any of the tests included in this handbook. The reader should acquire sufficient information to assist with the selection of a test or tests for assessing the language proficiency of students.

Some of the information included in this handbook relates to "technical standards." Technical standards are, "those guidelines related to an assessment that are specified by psychometricians, statisticians, test publishers, and specialists in the domains covered by the assessment." (Wheeler & Haertel, 1993, p.139). Usually technical standards include the reliability and validity tests conducted for any instrument. When considering the technical standards of a test, it is important to keep the characteristics of the children to be tested in mind. By characteristics we mean the ages, grades, language background, exposure to United States public school systems, including the contexts and values of these systems, and other factors that may have an impact on the students' ability to perform on the test. The validity of the test may be compromised if the students in your school or program are not represented in the norm group for the test.

In order to make the meaning of these standards clear and accessible to the reader, we also have included a glossary of terms in this section. If you are interested in additional information on the topic of testing and technical standards, we recommend the following useful references:

American Psychological Association. (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

Anastasi, A. (1988). *Psychological Testing* (sixth edition). New York, NY: MacMillan Publishing Company.

Durán, R.P. (1988). Validity and Language Skills Assessment: Non-English Background Students. In H. Wainer & H.I. Braun (Eds.), *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Wheeler, P. & Haertel, G.D. (1993). *Resource Handbook on Performance Assessment and Measurement: A Tool for Students, Practitioners, and Policymakers*. Berkeley, CA: The Owl Press.

There are many useful publications available on testing practice and ethics. As minority language students are sometimes not able to protect themselves adequately from unfair testing practices, we consider it very important to be careful, informed consumers of standardized language proficiency tests and other tests that might adversely impact a student's individual educational progress. The glossary of terms should provide the reader with background information necessary to understanding the descriptive information included in this handbook on the tests. If you are already familiar with these terms, please skip the glossary and begin with the test review sections.

Glossary

The following terms and definitions should prove useful in helping to understand the information to follow:

Basal and Ceiling Rules are guides for minimizing testing time. Test items are arranged in order of difficulty, with the easiest item first and the most difficult item last. The chronological age of the examinee is used to identify the basal or easiest items that examinee is expected to answer. The ceiling for a particular examinee is identified after a specific number of items are missed consecutively (anywhere from four items to an entire page of items may need to be answered incorrectly to identify the ceiling).

Correlation Coefficient is a statistical measure of the linear or curvilinear relationship between two variables, scores, or assessments. The correlation coefficient ranges from -1.0 to +1.0; when there is no relationship between two variables, it equals zero. A negative value indicates that as the value of one variable increases, the other variable tends to decrease. A positive value indicates that as one increases in value, so does the other or, that as one decreases in value, so does the other. Correlation is used in both reliability and validity studies for tests.

Criterion Referenced Test (CRT) is a test developed and used to estimate how much of the content and skills covered in a specific content area have been acquired by the examinee. Performance is judged in relation to a set of criteria rather than in comparison to the performance of other individuals tested with a norm-referenced test (NRT).

Discrete Point refers to test items that measure a single unit of content knowledge in a particular domain. These items are usually multiple choice, true-false, or fill-in-the-blank and allow for only one correct answer.

Holistic Score is the assignment of a single score that reflects an overall impression of performance on a measure. Scores are defined by prescribed descriptions of the levels of performance, examples of benchmarks at each level, or by scoring rubrics.

Interpolation is a process of using available points based on actual data to estimate values between two points. Interpolation is usually done by calculation based on the distance between two known points but also can be done geometrically by connecting the points.

Language Dominance refers to the general observation that a bilingual or multilingual individual has greater facility in one language as opposed to the others. However this linguistic facility can vary based on the

context of language use (e.g., school, church) and linguistic skill (speaking, writing).

Lexical refers to the lexicon of a language and is roughly equivalent to the vocabulary or dictionary of a language. Another term used by linguists at this level of a language is *semantics*. Semantics is the study of meanings at the word or sentence level.

Morphological refers to how words are constructed; a morpheme is essentially the smallest unit of language which conveys meaning. Some morphemes can stand alone (e.g., test) while others can only appear in conjunction with other morphemes (e.g., ed).

Norm Group is the sample of examinees drawn from a particular population and whose test scores are used as the foundation for the development of group norms. *Group norms* are the statistical data that represent the average (usually mean score) results for various groups rather than the scores of individuals within one of these groups or individuals across all groups. Only to the degree that persons in the norm group are like the persons to whom one wishes to administer a test can proper interpretation of test results be made. In other words, the test is not valid for persons who are not like the norm group.

Norm Referenced Test (NRT) is an instrument developed and used to estimate how the individuals being assessed compare to other individuals in terms of performance on the test. Individual performance is judged in comparison to other individuals tested, rather than against a set of criteria (criterion referenced test) or in a broad knowledge area (domain referenced test).

Normal Curve Equivalent (NCE) is a transformation of raw test scores to a scale with a mean of 50 and a standard deviation of 21.06. NCEs permit conversion of percentile ranks to a scale that has equal intervals of performance differences across the full range of scores and which can be arithmetically manipulated. Percentile ranks can not be used in arithmetic calculations.

Phonological (graphonic) refers to the smallest, distinguishable unit of sound in a language which help convey meaning (i.e., a phoneme). In isolation, however, the phoneme /p/ means nothing. Graphonic refers to the visual, graphic representation of the phonological system of a language which make reading and writing possible.

Pragmatic is the dimension of language which is concerned with the appropriate use of language in social contexts. In other words, pragmatics has to do with the variety of functions to which language is put (e.g., giving instructions, complaining, requesting) and how these functions are governed depending on the social context (e.g., speaking to a teacher versus a student; at school versus at home).

Reliability is the degree to which a test or assessment consistently measures whatever it measures. It is expressed numerically, usually as a correlation coefficient. There are several different types of reliability including:

Alternate Forms Reliability is sometimes referred to as parallel forms or equivalent forms reliability. Alternate forms of a test are test forms designed to measure the same content area using items that are different yet equivalent. This type of reliability is conducted by correlating the scores on two different forms of the same test.

Intra Rater Reliability is the degree to which a test yields consistent results over different administrations with the same individual performing at the same level by the same assessor

(intra-rater).

Inter Rater Reliability is the degree to which an instrument yields the same results for the same individual at the same time with more than one assessor (inter-rater).

Internal Consistency Reliability is sometimes called split half reliability and is the degree to which specific observations or items consistently measure the same attribute. It is measured in a variety of ways including Kuder Richardson 20 or 21, Coefficient Alpha, Cronbach's Alpha, and Spearman Brown Prophecy Formula. These methods yield a correlation coefficient that measures the degree of relationship between test items.

Rubric is sometimes referred to as a scoring rubric and is a set of rules, guidelines, or benchmarks at different levels of performance, or prescribed descriptors for use in quantifying measures of attributes and performance. Rubrics can be holistic, analytic or primary trait depending upon how discretely the defined behavior or performance is to be rated.

Stratified Sampling Design is the process of selecting a sample in such a way that identified subgroups in the population are represented in the sample in the same proportion that they exist in the population.

Syntactic level of a language is equivalent to the grammar of the language. This level (sometimes referred to as syntax) involves the way words are combined in rule-governed order.

Validity is the degree to which a test measures what it is supposed to measure. A test is not valid per se; it is valid for a particular purpose and for a particular group. Validity evidence can come from different sources such as theory, research or statistical analyses. There are different kinds of validity including:

Content Validity is the degree to which a test measures an intended content area. Item validity is concerned with whether the test items represent measurement in the intended content area. Sampling validity is concerned with how well the test samples the total content area. Content validity is usually determined by expert judgement of the appropriateness of the items to measure the specified content area.

Construct Validity is the degree to which a test measures an independent hypothetical construct. A construct is an intangible, unobserved trait, such as intelligence, which explains behavior. Validating a test of a construct involves testing hypotheses deduced from a theory concerning the construct.

Concurrent Validity is the degree to which the scores on a test are related to the scores on another, already established test administered at the same time, or to some other valid criterion available at the same time. The relationship method of determining concurrent validity involves determining the relationship between scores on the test and scores on some other established test or criterion. The discrimination method of establishing concurrent validity involves determining whether test scores can be used to discriminate between persons who possess a certain characteristic and those who do not, or those who possess it to a greater degree. This type of validity is sometimes referred to as criterion-related validity.

Predictive Validity is the degree to which a test can predict how well an individual will do in a future situation. It is determined by establishing the relationship between scores on the test and

some measure of success in the situation of interest. The test that is used to predict success is referred to as the predictor and the behavior that is predicted is the criterion.

Organization of the Test Information

Each test review consists of 10 points of information that include some very concrete items such as the purpose of the test, administration time, cost and scoring. In addition, the theoretical foundation for the test and the technical standards of reliability and validity are addressed. The final point of information provided directs the reader to critiques of the test.

The authors reviewed all available test materials provided by the publishers including actual test items, technical and administration manuals and scoring forms, guidelines and computer programs. Test critiques were not available for the Spanish forms of these tests. However, all critiques about the English forms of these tests were reviewed. The tests were administered to a small non-random sample of students when necessary to understand the administration procedures and scoring. Some of the descriptions of the tests vary as a result of the information available about the tests in the manuals and other documentation. For example, the Woodcock-Muñoz Language Survey includes a wider variety of validity studies than is the case for some of the other tests. The validity studies for this test were reported as described in the test manual. Some of the other tests conducted fewer validity studies but more reliability studies. In each case, we reported the information available in the test manuals and so, each of the test descriptions varies dependent upon the test documentation.

Although we wanted to make the theoretical foundation and the reliability and validity sections in this handbook parallel for each test, this was not possible. The test developers used the theoretical foundation and conducted the reliability and validity studies which they felt necessary to the construction of their instrument. All test developers presented some evidence to support both the reliability and validity of their instruments. The reader is urged to consult one of the sources cited earlier in this section (p. 18) for additional information about the minimum requirements for establishing test reliability and validity and compare these guidelines to the information presented in the technical standards manuals for any of these tests.

Finally, to facilitate the process of choosing a test, we provide the reader with a checklist to assist them with the selection of a Spanish language proficiency test. The items on the checklist parallel the 10 categories we used to describe each test. Information about the test purpose, its method of administration, the cost, the time it takes to administer the test and so forth should be considered and rated for each of the 4 language modalities (listening, speaking, reading, and writing). A column is available for an overall rating as well. This checklist can be used for rating any language proficiency test or assessment and need not be considered only for the 5 tests described in this handbook. Some tests will not assess all 4 language modalities in which case the rater will check / rate only the modality columns on the checklist appropriate for that test. Use the checklist as you read this handbook or to review the "examination" kit for any test. Most test publishers will allow a 30 day examination period free of charge. Simply contact the publisher using the telephone number included in Figure 1 (in the Summary chapter of this handbook) and ask them for a test examination kit.

Checklist for Spanish Language Proficiency Test Selection

Rate each item using the following criteria:

- 5 = meets need exceptionally well
- 4 = meets the need
- 3 = some problems but acceptable
- 2 = problems detract form overall utility
- 1 = major problems with the instrument
- NA = does not apply to this instrument

Some tests assess all four language modalities (reading, writing, listening, speaking) and some do not. Columns have been provided for each of the language proficiency modalities so that the instrument can be rated for each of these areas if desired. One column is available for an overall rating.

topic/item	rating	speaking	listening	reading	writing	comments
The purpose of the test is clear and meets our definition of language proficiency.
The test administration directions are specific and understandable.
The test administration guidelines can be used by teachers and teacher assistants.
The cost of the test is OK.
The test items seem to assess our definition of language proficiency.
The test items can be used to help the teacher design appropriate instruction for the individual student.
The test scores are useful for program placement.
The test scores can be used for evaluating the program.
There are multiple forms of the test so that it can be used for Pre/Post testing to evaluate the program.
The amount of time it takes to

administer the test is OK.						
The type of administration format (individual or group) is acceptable.
The theoretical foundation of the test fits our definition of Spanish language proficiency.
The test reflects the cultural knowledge and beliefs specific to the Spanish background of the children who will tested with it.
The test identifies the Spanish-speaking country or countries included in item development and norm group sampling.
There is adequate explanation of the theoretical foundation for the test.
The test offers adequate evidence of reliability.
The type of reliability evidence provided fits the design of the test.
The type of validity evidence is adequate.
The type of validity evidence provided makes sense with regard to the purpose(s) for which the test was designed.
Major problems with the test, as identified by test critiques, do not compromise the positive elements of the test.

[\(table of contents \)](#)

Basic Inventory of Natural Language (BINL)

- [Purpose](#)
- [Age and language groups](#)
- [Administration time](#)
- [Cost](#)
- [Test administration](#)
- [Items](#)
-
- [Scoring](#)
- [Test design and theoretical foundation for test](#)
- [Reliability and validity](#)
- [Test critique references](#)

Purpose

According to the *Technical Report* (1991), the BINL can be used as an indicator of oral language dominance and/or language proficiency. It also can be used as a tool to monitor language development. The BINL (Spanish) scoring system leads to the following four language proficiency classifications:

- Non-Spanish Speaking (NON)
- Limited Spanish Speaking (LTD)
- Functional Spanish Speaking (FNC)
- Proficient Spanish Speaking (PRO)

Age and Language Groups

According to the *Technical Report*, the BINL is appropriate for students in grades K through 12 between the ages of 4.9 and 17.8 years. The test was normed on a sample of 6529 students. The Technical Report provides Spanish Normative Data (i.e., percentile scores, raw scores and normal curve equivalent scores) for each grade level. Approximately 57% of this sample were kindergarten students between the ages of 4.9 and 5.8 years of age. No other information about the Spanish language norming sample is presented in the Technical Report.

Administration Time

The author of the BINL recommends that the test users take some time to familiarize themselves and the students with the test format (i.e., an oral picture describing type activity). Familiarization with the test can be done with small groups as part of regular classroom activity. The author estimates the average administration time to be about 10 minutes per student.

Cost

The cost of the complete BINL Kit for elementary (Grades K-6, Forms A or B) or secondary students (Grades 7-12, Forms C or D) is currently \$59.00. The kit includes: Examiner's Manual, 400 oral score sheets, class oral language profile card, 20 full-color poster story starters and 80 talk-tiles to elicit speech. The oral language samples can be machine scored through the publisher at a cost of 80 per score sheet. School districts may also opt for purchasing a rapid scoring program diskette for \$285. Other BINL products (e.g., instructional materials) and services (e.g., scoring written Spanish language samples) are also

available.

Test Administration

An oral language sample is elicited from the student using real-life, colored posters approximately 16" X 11" in size. The technique used to elicit the oral sample is referred to as a "free-sampling technique." Essentially, the intent of the test is to obtain as natural an oral sample from the student as possible within a picture-description type of test format. The student selects 3 to 5 pictures s/he prefers to talk about. The person administering the test must tape-record the student's oral sample and must avoid using direct questions to prompt the student or correct the student. The test administrator is asked to keep the student in an information-giving role and to maintain the testing situation as a logical communication activity. Logically, the test administrator must be proficient in the Spanish language to administer this form of the test.

Items

Because the BINL uses a free-sampling technique, there are no specific test items. However, the student's oral language sample is scored based on three linguistic features:

- fluency or the total number of words used in the language sample;
- level of syntactic complexity (i.e., grammatical competence); and
- average sentence length.

Scoring

The oral language sample can be either hand or computer scored. However, the author suggests computer-scoring when a large number of students are tested and staff are not trained in grammatical analysis. In either case, the scoring of the oral language sample hinges on the three factors previously mentioned: fluency, level of syntactic complexity and average sentence length. A fluency score or value is determined based on the total number of different words used in the language sample. The average sentence length value is generated based on the fluency count and the number of phrases or sentences used by the student.

The level of syntactic complexity score is calculated based on a predetermined numerical value assigned to various word classes. For example, in English, an adjective is assigned a value of 14, an adverb a value of 15, and a preposition a value of 20. The authors of the *Technical Report* (1991, p. 39) state:

The system, originally developed in English, was adapted to the scoring of the other languages in which the test is available. The scoring systems for English, Spanish, and Portuguese and other western languages are similar in that scores are assigned to nouns, verbs, auxiliary verbs, determiners, adjectives, adverbs, prepositions, and other grammatical devices common to the languages.

However, neither the Spanish language grammatical structures which are scored or their values are presented in the Technical Report. In short, the fluency, level of syntactic complexity, and average sentence length values are combined to generate one of the four possible Spanish language proficiency classifications.

Test Design and Theoretical Foundation for the Test

According to the author, the BINL is a criterion referenced instrument. The intent of the test is to obtain a natural, oral sample from the student within a picture-description type of test format. Unlike other tests, the BINL does not use specific test items to which students must respond and which are later scored as either correct or incorrect. Rather, the test design of the BINL is intended to measure language proficiency through the process of eliciting a natural language sample which is later examined for specific linguistic qualities.

The theoretical foundation underlying the BINL is best characterized as one which firmly rests upon the varying complexity of syntax or grammatical structures. In the *Technical Report* for the BINL (1991), proficiency is defined as "the degree of command of a particular language or dialect; that is, the ability a student has in using the various components of language such as vocabulary, structure and morphology" (p. 1).

The author of the BINL bases the design of the instrument on linguistic research which supports the following three findings. Vocabulary, syntactic (grammatical) structures and sentence length vary on a developmental continuum. In other words, research has shown that a young child's breadth of vocabulary, use of complex syntactical structures and ability to use longer and longer sentences develops gradually and predictably as the child grows older. However, it is the child's ability to use syntactic structures (i.e., various word classes, types of phrases and sentence types) of varying complexity that provides the theoretical foundation for the BINL.

Reliability and Validity

Evidence for the reliability of the BINL (Spanish) is based on one correlation analysis. The language classification of a student is based on what is called the complexity level score. This score is derived by scoring ten orally produced sentences. The authors of the test performed a split-half correlation. That is, the author examined the correlation between the complexity level score of the first five sentences with the complexity level score for the remaining five sentences. A correlation of .927 was found in this analysis involving 899 Spanish dominant students in grades K through 6 from San Diego, California.

No evidence of the validity of the BINL (Spanish) is provided in the *Technical Report* of this test.

Critiques

No critiques of the BINL (Spanish) could be found at the time of this publication. However, three critiques of the BINL as an English language proficiency test can be found in the following sources:

Carrasquillo, A. (1992). Basic Inventory of Natural Language. In D.J. Keyser & R.C. Sweetland (eds.), *Test Critiques Volume IX*. Austin: Pro-ed publications.

Guyette, T.W. (1985). Review of the Basic Inventory of Natural Language. In J.V. Mitchell (ed.), *The Ninth Mental Measurements Yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Hargett, G. (1987). Basic Inventory of Natural Language. In J.C. Alderson, K.J. Krahnke, & C.W. Stansfield (eds.), *Review of English Language Proficiency Tests*. Washington, D.C.: Teachers of English to Speakers of Other Languages.

[\(table of contents \)](#)

Bilingual Syntax Measure I and II (BSM I & II Spanish)

- [Purpose](#)
- [Age and language groups](#)
- [Administration time](#)
- [Cost](#)
- [Test administration](#)
- [Items](#)
- [Scoring](#)
- [Test design and theoretical foundation for test](#)
- [Reliability and validity](#)
- [Test critique references](#)

Purpose

Based on the *Technical Manual* of the Spanish version of the BSM I (Burt, Dulay, & Hernández-Chávez, 1976), the purpose for this test is to provide a measure of oral language proficiency. The *Technical Handbook* for the BSM II (Burt, Dulay, Hernández-Chávez, & Taleporos, 1980) indicates that this version of the test was designed for the same purpose as the BSM I. The target population for BSM I is students in grades K through 2 and the BSM II is for students in grades 3-12. Both tests lead to the following Spanish language proficiency classifications:

- Level 1: No Spanish
- Level 2: Receptive Spanish Only
- Level 3: Survival Spanish
- Level 4: Intermediate Spanish
- Level 5: Proficient Spanish I

The BSM II uses the same classifications as the BSM I from Level I to Level 5. The BSM II differs from BSM I in that two additional classifications are possible:

- Level 6(N): Proficient Spanish II
- Level 6 (S): Proficient Spanish II

The difference between Level 6 (N) and Level 6 (S) is that students scoring at Level 6 (S) display a greater use of standard syntactic structures than students scoring at Level 6 (N). The results from either of the two tests can be used to make placement decisions, provide formative and summative program evaluation data, make diagnoses, and collect language acquisition research data.

Age and Language Groups

According to the *Technical Handbook* for the BSM I designed for students in grades K-2, the test was field tested in 1974. Following the field test, 914 students were administered both the English and Spanish version of the test; an additional 232 pupils were tested only in Spanish. The authors of the BSM *Technical Handbook* (1976, p. 34) state, "The BSM sample should not be considered a random sample of all United States pupils enrolled in bilingual programs in the Spring of 1974." These students share some of the

following characteristics:

- 55.8 % of the sample learned Spanish as a first language and 36% learned English first; 7.7% learned Spanish and English as first languages.
- The age of the students ranged between 69.4 and 83.0 months.
- 33.8% of the Spanish speaking sample came from urban communities, 28.1% from rural communities, 17% from suburban communities and 21.1% came from an urban-rural mixed community.
- 12% of the Spanish speaking sample came from the North Atlantic region of the U.S., 10.8% from the Southeast, 8.9% from the Great Lakes & Plains, and 68.3% came from the West and Southwest.
- 47.6% of the total sample were classified as Chicano, 8.4% Cuban, 10.8% Puerto Rican, 2.9% Other Spanish, 5.1% Afro-American, 23.9% Anglo-American, and 1.3% were classified as Other non-Spanish.

The Spanish version of the BSM II, intended for learners in grades 3-12, was field tested in 1977 on 515 students sharing some of the following characteristics:

- 60.9% learned Spanish as a first language and 17.8% learned English as a first language; 9.3% learned an other first language, 6.3% did not respond and 5.7% learned Spanish and another language as a first language.
- The age of the students ranged between 8.7 and 17.9 years of age.
- The sample came from the same geographic regions as the BSM I; however, no percentages are provided;
- 60% of the sample were classified as Mexican-American, 12% other Hispanic, 1.8% Black, 17% White (non-Hispanic), 3.4% Other and 5.8% no response.
- The sample drawn was evenly divided by males and females.

Again, the authors of the *Technical Manual* (1980) for the BSM-II state that the sample was not intended to be a random sample of different language and socioeconomic groups.

Time

The administration time for the BSM I is approximately 10 to 15 minutes per child and may vary depending on how quickly the child responds, how long the responses are and how quickly the examiner can write the child's responses down. Neither the *Technical Handbook* (1980) nor the *Manual* (1978) for the BSM II indicate how long it takes to administer the BSM II. Hayes-Brown (1987) indicates it takes 10-15 minutes to administer the English version of the BSM II and Carrasquillo (1988) states from 15 to 20 minutes.

Cost

The current cost for a complete kit of the BSM I or BSM II, which includes the English and Spanish

editions of the test, is \$302.50. The kit includes a storage box, picture booklet, administration manuals in both English and Spanish, *Technical Manual*, 35 response booklets in both languages and 2 class record sheets. If only the English version of the BSM I or BSM II is needed, the above items may be purchased separately for a total cost of \$206.50.

Test Administration

The *Administration Manual* (1975) for the Spanish version of the BSM I is written in Spanish. The *Manual* describes the administration of the test in the following manner:

El procedimiento para suministrar la BSM-S consiste en sostener una conversacion con el nino acerca de algunos dibujos agradables sobre los cuales se le hacen una serie de preguntas. [The procedure for administering the BSM-S is like having a conversation with a child regarding some nice pictures about which a series of questions are asked] (p.6)

The tool used to elicit the oral language sample from the student is a series of colorful, cartoon-like pictures. There are a total of 25 test questions, some of which correspond only to specific pictures. Some questions also require the examiner to point to specific aspects of the pictures as the questions are asked. Further, these questions are designed to elicit the use of specific Spanish syntax structures. The examiner must write the student responses exactly as they are produced on the appropriate line in the test response booklet. The examiner must be able to speak, understand, read and write Spanish and maintain a natural flow of conversation throughout the administration of the test.

The *Administration Manual* (1975) for the Spanish version of the BSM II is written in Spanish. Because the BSM II is intended for older students, a few aspects of the test administration are different than the BSM I. For example, the cartoon story line for the BSM II is more sophisticated and contains a beginning, middle and end. In addition, the questions asked by the test administrator elicit more complex grammatical structures appropriate for older students. Nonetheless, the administration procedure for the BSM II is much like the procedure described for the BSM I. That is, the test administrator is asked to try to establish a smooth and real conversational exchange with the student.

Items

The items on the BSM I are the 18 (scored) test questions. Another 7 questions are asked but not scored. Some of the test items have underlined words which should be stressed. As stated previously, the test questions are designed to elicit the use of specific syntactic structures. For example, the BSM I is designed to elicit the use of prepositions, reflexives, linking verbs, articles, irregular verbs, and complex verb tense endings. It is also important to indicate that the test questions do not move from the use of simple syntactic structures to progressively more and more complex syntactic structures.

The test items on the BSM II consist of 20 (scored) questions also orally presented by the examiner and intended to elicit the use of specific syntactic structures appropriate to learners in grades 3 through 12. Additional test questions are asked but need not be scored. The grammatical constructions elicited by the questions entail structures such as complex sentence types, indirect questions, and conditional and subjunctive auxiliaries. Similar to the BSM I, the syntactic structures elicited vary in their complexity but are not necessarily elicited in an increasing order of difficulty.

Scoring

The BSM I and the BSM II are hand scored. The procedures for determining a student's proficiency level in Spanish require basically the same three steps. First, the scorer must make a determination as to whether or not a child responded to a minimum number of test questions in Spanish. If this criterion is met, then the scorer proceeds to step two. At this point, the scorer makes a determination as to the number of grammatically correct student responses to a given set of the test questions. Specific guidelines are provided for determining grammaticalness. Finally, the student's level of proficiency is determined based on the number of correct responses to specific test items. In effect, those students demonstrating competency with more advanced syntactic structures are classified as more orally proficient in the Spanish language.

Test Design and Theoretical Foundation for the Test

As its name suggests, the Bilingual Syntax Measure I & II are tests whose foundation rests on syntax or what is sometimes referred to as grammar. The *Technical Handbook* (1980, p. 3) for the BSM II states:

The Bilingual Syntax Measure series (both BSM I and BSM II) is built upon two fundamental concepts about language development:

- 1) Children acquire language largely through a process of "creative construction."
- 2) Language structures are acquired in an ordered, hierarchical fashion: certain grammatical structures are learned before other more complex structures.

Briefly, the Bilingual Syntax Measure I & II are based on research which demonstrates that children acquire syntactic structures in a systematic and hierarchical manner as they progress through stages in language acquisition. Further, the authors drew on research which has shown that even children from diverse language backgrounds appear to acquire certain syntactic structures in a common order.

There is one fundamental difference between the test design of the BSM I and the BSM II. According to the BSM II Technical Handbook (1980), the BSM II is based on *Item Hierarchies* as opposed to *Structure Hierarchies*. In other words, the raw data which underlies the design of the BSM II are based on the responses that field test students gave to pairs of items and not on a structural analysis of their responses.

Reliability and Validity

Two analyses were conducted in 1974 to demonstrate the reliability of the Spanish version of the BSM I. The first analysis was a Test-Retest Reliability procedure. In this analysis, 143 K-2 Spanish speaking students from two regions of California made up the sample. The students were tested twice, two weeks apart, in order to examine consistency in their language proficiency classification. The results of this procedure indicated that only one child was classified more than one level apart on the two administrations of the test.

The second procedure entailed a measure of inter-rater reliability. In this case, two scorers were given the same set of student responses to score and the student classifications were then compared for consistency. For this analysis, 242 tests from students in grades K-2 were scored by two independent scorers; 31 children were classified at a higher level by one of the scorers, while 17 students were classified at a lower level of

Spanish language proficiency. In short, the scorers agreed on the student classification approximately 80% of the time. The authors of the test attribute this discrepancy to the fact that a number of different dialects of Spanish were represented within their sample.

The reliability of the Spanish version of the BSM II included various analyses. An analysis of internal consistency yielded an overall coefficient of .82. However, the number of participating students and other relevant student data on which this analysis is based are not given.

Analyses of test-retest reliability were based on the scores of 80 students in grades 4 through 8 who took the test between one and two weeks apart. The results of this analysis indicate that 71 of the 80 children were classified at the same level and no children were classified at more than one level apart on the two testings. Overall, the inter-rater reliability for this version of the test is approximately .88.

The *Technical Manual* (1976, p. 32) of the BSM (I) sets forth three pieces of evidence which the test developers believe support the construct validity of both the Spanish and English versions of the test:

1. the body of psycho linguistic theory and research supporting the concept of natural sequences of acquisition;
2. the statistical procedures and research which have resulted in the production of the acquisition hierarchy and of the scoring system which is based on the hierarchy; and
3. evidence that the BSM classifications reflect the relationships expected to be found among bilingual children.

The *Technical Handbook* for the BSM II (1980) indicates that the techniques for the construct validation of this test were similar to those used for the BSM I. Again, the main difference lies in the fact that the BSM II is based upon what the authors call an *Item Hierarchy*. Briefly, the essence of this hierarchy is linked to how the field test students responded to pairs of items. The response patterns of the students allowed the test developers to ascertain which items were more difficult than others and consequently to more precisely determine their value within the scoring system.

Critiques

As of the writing of this handbook, no critiques specific to the Spanish versions of the BSM I or II were located. However, critiques of the English versions of the Bilingual Syntax Measure I and II can be found in:

Bejar, I. (1978). Review of the Bilingual Syntax Measure. In O.K. Buros (Ed.), *The Eighth Mental Measurements Yearbook*. Highland Park, NJ: Gryphon.

Carrasquillo, A. (1988). Bilingual Syntax Measure II Test. In D.J. Keyser and R.C. Sweetland (eds.), *Test Critiques Volume VII*. Kansas City, MO: Test Corporation of America.

Cziko, G. (1987). Bilingual Syntax Measure I. In J.C. Alderson, K.J. Krahnke, & C.W. Stansfield (eds), *Review of English Language Proficiency Tests*. Washington, D.C.: TESOL.

García, E. (1985). Review of the Bilingual Syntax Measure II. In J.V. Mitchell (Ed.), *The Ninth Mental*

Measurements Yearbook Volume I. Lincoln, NE: University of Nebraska Press.

[\(table of contents\)](#)

Spanish IDEA Proficiency Tests (IPT)

- [Purpose](#)
- [Age and language groups](#)
- [Administration time](#)
- [Cost](#)
- [Test Administration](#)
- [Items](#)
- [Scoring](#)
- [Test design and theoretical foundation for test](#)
- [Reliability and validity](#)
- [Test critique references](#)

Purpose

Educators and evaluators had a need for the comparison of students' home language proficiency with their English language proficiency and this need prompted the development of the Spanish IDEA Oral Language Proficiency Test. The English IPT was designed to assist districts with the process of identifying limited English proficient students (LEP) and with redesignating these students as fluent English proficient for placement in mainstream classes after a period of instruction in special programs for LEP students. The Spanish Oral IPT was developed by Enrique Dalton as a parallel version of the English test. The Spanish IPT scoring system leads to the following categories of oral language proficiency:

Spanish Oral IPT Language Proficiency Designations	Spanish IPT Reading Designations	Spanish IPT Writing Designations
Non-Spanish Speaking	Non-Spanish Reader	Non-Spanish Writer
Limited Spanish Speaking	Limited Spanish Reader	Limited Spanish Writer
Fluent Spanish Speaking	Competent Spanish Reader	Competent Spanish Writer

The Spanish parallel version of the Oral IPT was field tested in the spring of 1980 and was published later in the same year. The IPT Spanish Reading and Writing test, levels 1, 2, and 3, was developed in 1993 and pilot tested in the winter of 1994. These tests were published in the summer of 1994. Together the Spanish IPT Oral and the Reading and Writing Tests assess overall Spanish language proficiency.

Age and Language Groups

Several different versions of the Spanish IPT were developed for specific age groups. The Pre-IPT Spanish was designed to test the oral language proficiency of preschool students. The Spanish IPT was designed to

assess oral proficiency for students in grades K - 8. The Spanish Reading and Writing Tests (IPT-Spanish Reading & Writing Test) are sold in three versions: Level 1 is for grades 2 - 3; Level 2 is for grades 4 - 6; and Level 3 is for grades 7 - 12.

Each of the versions of the test has information about the configuration of the norm group sample including names of schools, districts, and states participating, the age, gender, and ethnicity of the students in the norm group, and the primary languages of the students. As there are several different versions of the Reading and Writing Test and two different versions of the Spanish Oral IPT, the information about the norm group for the Levels 1 and 2 Spanish Reading and Writing Test is presented for illustrative purposes (Dalton, Barrett, & Jensen, 1994, pp. 7-9). The norm group was characterized by:

- 50.0 % of the sample was male and 50.0% of the sample was female;
- states included in the sample were CA, FL, IL, KS, NM, NY, TX, and VA
- student ages ranged from 7 years (11.3% of the sample) to 13 years (3.3% of the sample). The percentage of students for each age between 7 and 13 years was reported;
- the ethnic groups included in the sample were: Hispanic (98.2%), Caucasian (1.5%), and Black (0.3%);
- students in second grade comprised 19.7% of the sample, third grade 20.8%, fourth grade 18.5%, fifth grade 23.2%, and sixth grade, 17.8%.

The reader is urged to review the norm group information for the specific form of the Spanish IPT to be used.

Administration Time:

The average testing time is 14 minutes. The time it takes to administer the test will vary depending upon the promptness of the student's responses. Students often will not need to complete the test or will not need to start at the beginning of the test so the testing time will be very short.

The Spanish Reading and Writing Tests are untimed. The *Examiner's Manual* for the various forms provides a table of approximate time ranges for each section of the reading/writing tests. The Reading Test will take from 55-80 minutes to administer (Amori & Dalton, 1994, p. 8).

Reading	
Part 1: Vocabulary	7-12 minutes
Part 2: Vocabulary in Context	7-12 minutes
Part 3: Reading for Understanding	15-23 minutes

Part 4: Reading for Life Skills	10-15 minutes
Part 5: Language Usage	6-10 minutes

The Writing Test will take from 30-55 minutes to administer (Amori & Dalton, 1994, p. 8).

Writing	
Part 1: Conventions	5-10 minutes
Part 2: Write a Story	6-15 minutes
Part 3: Write Your Own Story	15-20 minutes

The IPT publisher recommends allowing a break between the Reading and Writing Tests. The Writing test can be given later in the day or on the next day.

Cost:

The cost of the Oral Proficiency Test set with 50 test booklets is \$105.00. An additional 50 test booklets is \$29.00. Components of the Oral Language Test set can be purchased separately as well. The Reading and Writing Test Set prices vary. For the IPT Reading & Writing Tests 2 and 3 (grades 4-6 and 7-12 respectively), each set includes 50 reading test booklets, 50 reading test answer sheets, 50 writing test booklets (consumable), a scoring template, and an *Examiner's Manual* and a *Technical Manual* for \$160.00. The IPT Reading & Writing Test 1 set (for grades 2 and 3) includes the same components except that there are 50 consumable reading test booklets and no test answer sheets (students write their answers in the test booklets) for \$150.00. Extra reading test answer sheets are \$25.00 and extra test writing test booklets are \$37.00. The technical and examiner's manuals are \$29.00 and \$39.00 respectively.

Test Administration

The Spanish IDEA Oral Language Proficiency Test assesses four basic areas of Spanish oral language proficiency: Vocabulary, Comprehension, Syntax, and Verbal Expression which includes Articulation. There are six levels of difficulty tested (A - F) and all students are tested individually. A student advances through the levels of the test until the test is completed or stops at a proficiency level as indicated by information in the score box at the end of each level.

The Spanish IPT Reading and Writing Test is a group administered, standardized test. The Reading Test consists of five parts: Vocabulary, Vocabulary in Context, Reading for Understanding, Reading for Life Skills, and Language Usage. The Writing Test has three parts: Conventions, Write a Story and Write Your Own Story. Three different reading level designations (a non-Spanish reader, a limited Spanish reader, and a competent Spanish reader) and three writing level designations (a non-Spanish writer, a limited Spanish writer and a competent Spanish writer) are identified after the Reading and Writing Test is scored.

Items

In the appendix of the *Examiner's Manual* for the IPT I Oral Proficiency Test is a matrix of test items for each skill area and developmental level. The oral proficiency items consist of questions and statements to the student (e.g., "Stand up and turn around", or "What is your name?") and colorful picture cards for identification of an action verb (e.g., riding a bike, were singing/sang) or a noun (e.g., helicopter, stove).

The items for the Reading and Writing Test are presented in a traditional multiple choice, bubble in the correct answer format. Reading test items have four possible responses among which to choose. The writing multiple choice items have three possible answers. The Writing test includes two sections that require the student to write a story. The first section has a series of three pictures that "tell" a story. The student is expected to write about the sequence of events depicted. The second writing sample has two different picture prompts and the student is expected to write a story about one of the prompts. Other forms of the Reading and Writing Tests include similar items that are appropriate developmentally for the students' age. The test consumer needs to keep in mind the educational experiences and background of the examinee for the Reading and Writing test to be valid.

Scoring

Scoring the Spanish Oral Language Proficiency Test is a straightforward procedure. Each item has a corresponding response line which gives the desired answer. It is numbered the same as the item. Each item is followed by two boxes: one to check if the item is correct and the other to check if the item is incorrect. The Oral Test is divided into six levels, A-F. At the end of each level, the examiner totals the number or errors or incorrect responses. Each level gives a maximum number of errors possible to continue the testing at the next level. If the student exceeds an error threshold, they are classified at that level or the previous level depending upon the total number of errors made. These six levels can then be translated in the NES, LES, or FES categories.

The multiple choice portions of the Reading and Writing Tests can be hand scored using a template or they can be machine scored by the publisher. The total score for this test is the total number of correct answers in each of the five subtest sections. The number correct then can be compared with the required number of correct test items for the Non-Spanish reader, Limited Spanish reader and Competent Spanish Reader designations. A table in the Appendix of the *Examiner's Manual* can be used to interpret the raw scores as standard scores, percentile ranks, or normal curve equivalents (NCE).

Scoring the Writing Test involves a similar process for the multiple choice items but the two sections that involve actual student writing samples completed for the test booklet prompts, require the examiner to use the IPT Rubrics and student examples in the *Examiner's Manual* to rate the writing samples. The rating given to these two sections using the rubric is considered the score. The scores for the three writing sections are compared to a range of scores that are required for classification as a non-Spanish writer, a Limited

Spanish Writer or a Competent Spanish writer. The writing test can be machine scored; however, the writing samples must be rated by the examiner. Care must be taken to train examiners to use the rubric so that subjectivity and rater bias are not introduced into the scoring process. To ensure the reliability of these ratings, two examiners should rate each writing sample. If there is a discrepancy, a third rater should score the paper independently. Training scorers adequately using clear benchmark papers (samples are provided in the manual) will ensure inter-rater reliability.

Test Design and Theoretical Foundation for the Test

The Spanish IPT is made up of a series of reading, writing and oral language proficiency tests designed for use with students from pre-school through adult. The following Spanish and English forms of the IDEA IPT language proficiency tests are available:

Oral Language Proficiency Tests
Pre-IPT English (for preschool students)
Pre-IPT Spanish (for preschool students)
IPT English forms A, B, C, D (K- grade 8)
IPT Spanish form (K- grade 8)
IPT II English forms A and B (grade 8 - adult)
IPT II Spanish (grade 8 - adult)
Reading and Writing Proficiency Tests
IPT -1 English (grades 2-3) forms 1A and 1B
IPT -2 English (grades 4-6) forms 2A and 2B
IPT -3 English (grades 7-12) forms 3A and 3B
IPT -1 Spanish (grades 2-3)
IPT -2 Spanish (grades 4-6)
IPT -3 Spanish (grades 7-12)

Six basic beliefs or premises comprise the underlying structure for the IPT family of tests: 1) language is developmental; 2) language is incremental; 3) language is systematic; 4) language is symbolic; 5) language is used to communicate in a social context; and 6) language involves both receptive and productive skills. The theoretical foundation for the IPT series is based on the four stages of language acquisition: babbling; echolalic; telegraphic; and syntactic. The babbling stage is from birth to 12 months of age and is the period during which infants play with sounds such as gurgling, cooing, raising and lowering pitch, laughing and crying. These sounds are one way the infant communicates with other people in the environment. Meaning

is not attached to infant sounds until infants accidentally articulate sounds to which parents attach meaning (e.g., mammaa). At this point parents tend to repeat the sounds, smile at the child, and reinforce the sounds in a variety of other social ways. The child is then more likely to repeat the sounds.

In the echolalic stage, the child begins to use inflections, stops, pitch, pauses, and stresses in imitation of the sound patterns of the language the child hears spoken. It is during this stage that the child acquires and practices the various phonemes (individual sound units) and morphemes (syllables) of the language. During this stage, the child begins to connect a series or combination of sounds with the objects in the environment for which they stand. Labeling begins and the child uses these labels to name and describe the things around in the environment and to express needs and feelings. This stage is called the telegraphic stage because the child uses one and two word combinations to convey meaning.

The final stage is the syntactic stage. During this stage the child begin to assimilate the rules or grammar of the language. At first, simple two and three word sentences are constructed. Later as the child acquires new adjectives, adverbs, prepositions, and verbs, more complex communication becomes possible. Further interaction with proficient adult speakers allows the child to refine language use and communicate more clearly with those important people in the world around.

Reliability and Validity

Reliability of the Spanish Oral Proficiency Test (I & II) and the Pre-IPT Spanish is extremely high. Reported internal consistency reliability, based on inter item correlations, was .99 which is unusually high. Split-half reliability for the Oral Test was also high with a .99 coefficient reported (Dalton, 1980, p. 20).

The reliability for the Reading and Writing Test was reported for each of the different subtests and forms. The reliability coefficients quoted in this section are for IPT 1 & 2 (grades 2-6) Spanish Reading and Writing tests. The reliability analyses for the Reading subtests (for the Level 1 test) are reported first and include the conventions Writing subtest. Reliability was established in a different manner for the other two Writing subtests separately from the Reading and Writing conventions reliability coefficients. Internal consistency was measured for each of the Reading subtests and the Writing Conventions subtest and ranged from .68 for the Writing Conventions subtest to .83 for the Reading for Understanding subtest. Internal consistency for the entire Reading battery was .94. These coefficients indicate a moderate to high degree of internal consistency. Test-retest reliability was measured also and ranged from .72 for Writing Conventions to .78 for Language Usage for the Level 1 test. The test-retest reliability for the entire battery was .85 which indicates that the results of comparing scores taken two weeks apart are very consistent.

Reliability for the Writing subtests involved establishing inter-rater correspondence between two different raters of the examinee's writing samples. Agreement on the reader rating for each of the four "items" of the two writing subtests that involved student writing samples was provided as a percentage. Exact agreement occurred 70% of the time for the first item, 61% for the second item, 59% for the third item, and 65% for the fourth item.

Validity for the Oral Language Test included discussion of content validity and criterion (concurrent) validity. For content validity, the reader is provided with a chart that displays the relationship between each test item and the six behavior domains measured by the test (syntax, lexicon, phonology, morphology, comprehension, and oral expression). The percentage of items devoted to measuring each domain is reported at the end of the chart and this information is offered as evidence of content validity.

Criterion validity (sometimes called concurrent validity) was tested by comparing teacher predictions of student performance with IPT results. Teachers' perception (opinion) of student language proficiency and IPT results was strongly related for the Spanish Oral IPT (the IPT II technical standards are reported separately in another technical manual) and the results of these analyses are described in a large table (p. 16). District designation of students as NSS/LSS/FSS, as measured using methods other than the IPT, are described in a table as well (p.17). Overall, many different methods and sources for evidence of the validity of the Oral IPT were presented and the reader is urged to consult the *Technical Manual* for the specific form of the IPT to be used (or purchased) and to review the reliability and validity evidence presented.

Validity for the Reading and Writing tests included the same types and data presentation format as was described for the Oral IPT. Construct, content, and criterion validity tests were conducted. A matrix of the items included in the Vocabulary, Vocabulary in Context, Reading for Understanding, Reading for Life Skills, and Language Usage subtests is organized by the objective competency or concept that each item tests. This matrix is provided to the test user for both the Level 1 and Level 2 tests as evidence of the content validity of the Reading and Writing Tests.

The subtest inter-correlations are provided in a table as evidence of the construct validity of the Reading and Writing Tests. Correlation coefficients for the Level 1 test range from a low of .44 for the correlation between the Vocabulary and Writing Conventions subtests to a high of .87 between the Total Reading Battery and Reading for Understanding subtest. The correlations in this table show the relationship between the various subtests. There are moderate correlations between the Reading subtests indicating that they are measuring related but not identical domains. The Reading subtests correlate higher among themselves than they do with the Writing Conventions subtest which is anticipated, as the writing domain is somewhat different from the reading domain.

Finally, criterion validity is presented for the Reading and Writing Test. It indicates how strongly the Spanish IPT Reading and Writing tests correlate with other independent measures of what the tests are designed to assess. Teacher ratings of students' reading and writing ability was used as an independent source of evidence for the purpose of comparison to IPT Reading and Writing Tests. When teacher ratings were compared to IPT Reading and Writing test scores, the correlation coefficients ranged from .33 to .59 for the Level 1 tests.

Critiques

Critiques were not available for the Spanish version of the IPT. Critiques for the English versions of the IDEA Proficiency tests can be found in the following documents:

McCollum, P. (1983). The IDEA Oral Language Proficiency Test: A Critical Review. in S.S. Seidner (Ed.). *Issues in Language Assessment, Volume 2*. Chicago, IL: Illinois State Board of Education.

McCollum, P. (1987). IDEA Oral Language Proficiency Test. in J.C. Alderson, K.J. Krahnke, & C. W. Stansfield, (Eds.). *Review of English Language Proficiency Tests*. Washington, D.C.: Teachers of English to Speakers of Other Languages.

Rivera, C., & Zehler, A. (1987). IDEA Proficiency Test II. in J.C. Alderson, K.J. Krahnke, & C. W. Stansfield, (Eds.). *Review of English Language Proficiency Tests*. Washington, D.C.: Teachers of English to Speakers of Other Languages.

Stansfield, C.W. (1992). IDEA Oral Language Proficiency Test. in D.J. Keyser & R.C. Sweetland (Eds.). *Test Critiques: Volume IX*. Austin, TX: Pro-Ed.

Stansfield, C.W. (1992). IDEA Oral Language Proficiency Test-II . in D.J. Keyser & R.C. Sweetland (Eds.). *Test Critiques: Volume IX*. Austin, TX: Pro-Ed.

[\(table of contents \)](#)

Language Assessment Scales (LAS)

- [Purpose](#)
- [Age and language groups](#)
- [Administration time](#)
- [Cost](#)
- [Test administration](#)
- [Items](#)
- [Scoring](#)
- [Test design and theoretical foundation for test](#)
- [Reliability and validity](#)
- [Test critique references](#)

Purpose

Though available for grades Pre-K-12, only the most recent forms of the LAS Spanish tests for grades 1-6 are described. These include: LAS-Oral (Form 1B) and LAS-Lectura/Escritura (Forms 1A & 2A). Collectively, the LAS consists of an oral, reading (i.e., Lectura) and writing (i.e., Escritura) language proficiency assessment system. According to the LAS Preview Materials Booklet (1991), LAS results may serve several purposes including: assessing the learner's language proficiency, placement decisions, and reclassification. Because the three parts of the test can be combined in different ways, at least three proficiency classifications are possible: a LAS-Oral Score, A LAS-Lectura and Escritura Score, and a Language Proficiency Index (LPI) which combines the Lectura/Escritura overall score (1 to 3) with the oral score (1 to 5). The proficiency levels are:

LAS-Oral Score	LAS Lectura/Escritura Score
5 = Hablante fluido del espanol	3 = Competente
4 = Hablante fluido del espanol	
3 = Hablante limitado del espanol	2 = Limitado

2 = Hablante limitado del español	
1 = No-hablante del español	1 = Bajo

Language Assessment Scales Language Proficiency Index Classifications

LPI (LE/O)	Category	Description
1/2 2/3	LSPa	low level reading and writing skills; mid level ("limited")listening and speaking skills
1/4 1/5	LSPb	low level reading and writing skills; high level ("proficient")listening and speaking skills
2/2 2/3	LSPc	mid level reading and writing skills; mid level ("limited")listening and speaking skills
2/4 2/5	LSPd	mid level reading and writing skills; high level ("proficient")listening and speaking skills
3/2 3/3	LSPe	high level reading and writing skills; mid level ("limited")listening and speaking skills
3/4 3/5	FSP	high level reading and writing skills; high level ("proficient") listening and speaking skills

Age and Language Groups

According to the Manual de administración y calificación (1994), the LAS-Oral (Form 1B) is intended for

use with students in grades 1-6. Form 1A of the LAS-Lectura/Escritura is designed for students in grades 2-3, while Form 2A is appropriate for learners in grades 4-6.

The LAS-Oral (Form 1B) was normed on 1264 students sharing some of the following characteristics:

- Students came from one of three sites: Matamoros, Mexico, San Francisco, California and San Antonio, Texas.
- The sample represented Spanish speaking students with little or no exposure to English; inversely, the sample consisted of English speaking Hispanic learners with little or no exposure to Spanish; however, some of the students had a bilingual background.
- Of the students tested, 69.14% were born in Mexico, 14.13% were born in the US, and the remaining came from various Spanish speaking countries such as Chile, Guatemala, Puerto Rico and Spain.
- 481 (38.05%) of the students came from a Spanish language background; 783 (61.95%) of the students came from an English language background.
- 236 (18%) of the students between grades 1-6 had a Spanish language background; 555 (44%) of the students in the same grade had an English language background.

The technical manual for the Lectura/Escritura portions of the LAS is forthcoming. An unpublished draft (De Avila and Duncan, 1995) of this manual provided the following information regarding the norming sample for these parts of the test.

- A total of 2177 students were tested for all three levels of the test; 828 students in grades 1-4 comprised the sample for Form 1A; 498 students in grades 2-6 made up the sample for Form 2A.
- The students came from over 10 Latin American Countries but primarily from Mexico, Puerto Rico and El Salvador.
- Approximately 14% of the entire sample had a significant portion (80% of total years of schooling) of their education in Spanish.

Administration Time

According to the *Oral Administration Manual* (1990), there are no specific time limits for administering this part of the test. Generally, students should be given enough time to respond to each item. It should be noted that the oral part of the test can be administered in its long form (i.e., Vocabulario, Vamos a escuchar, Cuentos, Sonidos mínimos en pares, and Fonemas) or its short form which only includes Vocabulario, Vamos a escuchar, and Cuentos. Obviously, less time will be required to administer the short form of the oral test.

Form 1A of the LAS-Lectura/Escritura tests may be group administered and may be timed or untimed activities. Duncan and De Avila (1994) indicate that time administration limits will vary depending on the competency of the student and that less proficient students should be allowed all the time needed to complete the various portions of the test. Nonetheless, the authors do set forth some basic guidelines for allocating time for this portion of the LAS system. The authors suggest that both tests can be administered

in one hour, but recommend that for less proficient students two 45 or 55 minute sessions be used. They suggest that the first session entail the administration of the following four subtests: Vocabulario, Fluidez, Lectura para información, Mecánica y expresión; the second session could be given over to the remaining two subtests, Completando oraciones and Qué está pasando?.

The authors make essentially the same recommendations for the administration of Lectura/Escritura Form 2A for Grades 4-6. The only difference is that for this form of the test students must complete three subtests (instead of two) in the second testing session, if two testing sessions are going to be used.

Cost

Currently the cost of the Language Assessment Scales-Oral Level I, Form 1B (Grades 1-6) examiner's kit is \$100.00. The kit includes: an *Administration Manual*, a *Scoring and Interpretation Manual*, a cassette with a story and items, cassette with pronunciation subtest items, a cue picture booklet, and a plastic holder for six audio cassette tapes. 50 individual consumable answer documents must be purchased separately at a cost of \$26.50. 50 student profile sheets can be purchased for a price of \$15.00. Administration and scoring manuals are available in Spanish and English.

Testing materials for the Lectura and Escritura parts of the Language Assessment Scales are not available in a kit form. For Spanish Form 1A (Grades 2-3) 35 consumable reading test booklets at a price of \$66 and 35 consumable writing test booklets at a price of \$32.25 must be purchased. For testing reading and writing with Form 2A (Grades 4-6), the following materials are needed and must be purchased at the indicated prices: 35 reusable reading test booklets are available at a cost of \$42.00; 35 consumable writing test booklets cost \$32.25; and 50 student answer documents can be purchased at a cost of \$26.50.

The reader should note that the *Examiner's Manual* for the Lectura/Escritura tests is only included if the reading tests are purchased; otherwise, the writing manual must be purchased separately at a cost of \$9.75. The readers should note that the Technical Manual for the Lectura/Escritura tests is in press.

Test Administration

The LAS-Oral (Form 1B) must be individually administered. The authors recommend that the test administrator (1) rehearse administering the test with family or friends, (2) read all materials, and (3) ensure that the students understand what they are expected to do before beginning each subtest. Further, the test giver should be a literate and fluent, proficient speaker of Spanish and able to discriminate between correct and incorrect responses. The general instructions for the test may be provided in whatever language, combination thereof or dialect. Naturally, the oral part of the test should be administered in a quiet area.

The first oral subtest, within the Vocabulario subtests, is *Cómo se llama?*. It consists of 10 pictorially represented items (i.e., nouns) which the test giver must point to and say *Cómo se llama?* and which the student must orally identify. The second Vocabulario subtest, *Qué están haciendo?*, also requires the use of 10 picture cues aimed at eliciting the student's ability to produce the Spanish gerund form (i.e., -ando and -iendo) of 10 frequently used verbs.

The *Vamos a escuchar* subtest consists of a tape-recorded dialog to which the student must first listen and then respond to 10 yes-no tape-recorded questions. One picture cue is also provided. Pausing the tape between questions is permitted. The *Cuentos* subtest requires the learner to first listen to a tape-recorded

story and then retell the story with the support of four sequential picture cues.

The two optional subtests, *Sonidos mínimos en pares* and *Fonemas*, are also tape-mediated. In the *Sonidos mínimos en pares* subtest the student hears 30 word pairs and must determine whether they sound the same or different. The *Fonemas* subtest requires the test taker to listen to and repeat a total of 30 words, phrases or sentences containing a target phoneme.

Again, both levels of the *Lectura* and *Escritura* tests may be administered in a small group setting but each student must generate his/her own answers in the provided test booklets. Proctors are recommended for groups of more than 15 students. The test administrator must follow specific instructions including reading instructions aloud to the students; instructions may also be provided in the student's native language. In general, sample items are provided to ensure that students are familiar with the task and marking their answers in an appropriate manner in the answer booklet before attempting each subtest. The test administrator is advised to follow the same recommendations mentioned in reference to the administration of the oral portions of the test.

Items

According to the *Manual de Administración Español*, Form 1B (1990) the ten items making up the *Cómo se llama?* subtest are common Spanish nouns used by monolingual children in a public school setting. The remaining 10 items on the *Qué están haciendo?* subtest are also described as commonly used gerund forms used by students in grades 1 to 6.

The 10 yes-no questions comprising the *Vamos a escuchar* subtest are intended to measure the learner's ability to understand daily conversations. The questions are asked in various tenses including the present, past and future tenses.

The *Cuentos* has no specific test items. However, the learner's story retell is examined from various perspectives including: vocabulary, syntax, organization, transitional words, and fluency.

Again, and as concerns the *Sonidos mínimos en pares* subtest, the *Manual de Administración español*, Form 1B (1990) indicates that these 30 test items provide information about the learner's ability to understand words and sentences in daily and academic conversations. Finally, the 30 *Fonemas* subtest items are intended to provide insight into the learner's ability to make himself understood in the Spanish language.

The test items on the *Lectura* and *Escritura* Form 1A subtests can be described in the following manner. The *Vocabulario* subtest consists of 10 multiple choice items each of which is pictorially represented and then followed by four printed words. These items focus on nouns and adjectives. The *Fluidez* subtest consists of 10 individually presented sentences containing a blank. Each sentence is followed by 4 options from which the student must select the most appropriate response or missing word. These items focus mainly on verbs and adjectives.

The test items on the *Lectura para información* subtest are 10 statements which follow a short reading passage. After reading the short passage, the student must answer True or False to each statement. The items require factual recall. The *Mecánica* and *expresión* subtest consists of 15 multiple choice items. Each item contains a blank and is followed by three options from which the student must choose. Items are aimed at punctuation and grammar.

In the Completando oraciones subtest the student is presented with 5 incomplete sentences. The student must write a grammatically acceptable ending for each sentence. The items are aimed at the student's ability to produce dependent and independent clauses. In the final subtest for Form 1A, *Qué está pasando?*, there are five picture-cued items. The items are designed to require the student to produce written sentences which describe what is happening in each picture. These items are designed to elicit contextually appropriate responses written in a grammatically acceptable form.

Form 2A uses the same Lectura and Escritura subtest formats as Form 1A. Form 2A, as noted earlier, has one additional subtest, *Vamos a escribir*. This subtest is designed to generate an essay. The student writes guided by a series of four sequenced picture-cues and a lead in first sentence. The essay is scored for vocabulary, organization, fluency, transitions and mechanics.

It is also important to indicate that De Avila and Duncan (1995) adhered to the following criteria in the selection of Lectura/Escritura test items:

- Items were selected according to the extent that they were representative of the Spanish language, particularly if the item content was different from English and specific to Spanish (i.e., exclamation point).
- Items were selected only if they showed a positive correlation with the Test-Total score and demonstrated a point bi-serial correlation of at least $+0.150$.
- Each item had to be passed by at least 80% of the Spanish educated students taking the test.

Scoring

The oral portions of the LAS Oral (Form 1B) use a discrete point scoring format, with the exception of the *Cuentos* subtest which is holistically scored. The Answer Booklet used by the test administrator contains the correct responses for the *Vocabulario*, *Vamos a escuchar*, *Sonidos mínimos en pares* and *Fonemas* subtests. As the test administrator proceeds from item to item, s/he must darken in the "bubble" for incorrect responses only.

The *Cuentos* subtest is holistically scored. The examiner must first write down the student's oral story retell as accurately as possible. Tape recording the retell is advised. The student's retell is then scored using a pre-established scoring rubric on a scale of 0 to 5. Note that the examiner must first undergo what the authors call a reliability exercise in order to ensure consistency in scoring this part of the test. The authors also indicate that the scorer must be a proficient, literate speaker of the Spanish language.

Once the incorrect items have been tallied and a holistic rating has been assigned to the oral retell, the test administrator can determine the student's total score, language proficiency level and category. Each of the five parts of the oral test are weighted and simple calculations are required in order to generate the final scores.

The *Cuentos* score accounts for 50% of the student's total oral score. The scoring manual also includes instructions for dealing with scores falling within the "gray areas". It is important to keep in mind that if the short form of the LAS-Oral is administered, the student's score is based on only the *Vocabulario*, *Cuentos* and *Vamos a escuchar* subtests. The long form would include the *Fonemas* and *Sonidos mínimos en pares* subtests. In either case, the *Cuentos* subtest accounts for one half of the student's final oral score.

The Vocabulario, Fluidez, Mecánica y expresión and Lectura para información reading subtests for grades 2-3 (Form 1A) are scored as correct or incorrect according to the Answer Key. The Completando oraciones and Qué está pasando? writing subtests are holistically scored according to a rating scale with values from 0 to 3. The Vamos a escribir subtest, which applies only to grades 4-6 (Form 2A), is holistically scored using a rubric that ranges from 0 to 5. As in the holistic scoring of the Cuentos, scorers of writing subtests must participate in the reliability exercises and be proficient, literate speakers of Spanish.

Once each of the reading and writing subtests have been scored, various scores can be produced including standard scores, competency levels and reading and writing categories. Recall that these reading and writing scores can also be combined with the LAS-Oral score to generate the overall Language Proficiency Index previously referred to in Section A. Purposes.

Test Design and Theoretical Foundation for the Test

According to De Avila and Duncan, the theoretical foundation underlying the LAS-Oral (Form 1B) is the same as the theoretical foundation underlying the English and earlier Spanish versions of the test (See Section 9 of the Oral Technical Report, 1990). The authors describe the essence of the theoretical model of language proficiency which guided the development of the oral portions of the test in the following manner:

The development of the LAS was based on a view of language as consisting of four linguistic aspects: phonology (phonemes, stress, rhythm and intonation), the 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) (De Avila & Duncan, 1982, p. 8).

The authors also maintain that the LAS-Oral represents a convergent approach to language assessment. That is, while each of the above linguistic subsystems are assessed independently and weighted differently, "it is the combined total score which is of ultimate critical importance for identification and programmatic treatment" (De Avila & Duncan, 1982, p. 8).

According to De Avila and Duncan (1995), the theoretical design underlying the Lectura/Escritura portions of the LAS "is based on exactly the same principles and assumptions that guided the development of the original English versions" (p. 2). Note that the LAS Reading and Writing tests in English represent a convergent approach to language assessment (Duncan & De Avila, 1988); total scores are derived from various content areas and scoring procedures (i.e., multiple-choice and holistic). However, the authors maintain that the test design of the English versions was guided by the review of various state guidelines, curriculum guides, expected learning outcomes and the scope and sequence of many commonly used ESL instructional programs. The authors do not indicate if the same is true for Lectura/Escritura Form 1A and Form 2A of the LAS.

Finally, De Avila and Duncan (1995) set forth one important assumption which underlies both the English and Spanish versions of the LAS family of tests. The authors hypothesize that there is linear relationship between language proficiency and academic achievement. That is, as a learner's language proficiency increases as measured by the LAS so does the learner's academic achievement.

Reliability and Validity

Based on the *Oral Technical Report* (De Avila & Duncan, 1990), analyses of internal consistency yielded the following reliability correlation coefficients. The overall reliability of the Sonidos mínimos en pares subtest is .892; the reliability coefficient for the Fonemas subtest is .861. The Vocabulario subtest which includes the *Cómo se llama?* and *Qué están haciendo?* subtests yielded a correlation coefficient of .957, and the item analyses on the *Vamos a escuchar* subtest generated a coefficient of .886.

Because the Cuentos subtest is holistically scored, inter-rater reliability analyses were conducted. These analyses were based on a random sample of 46 student responses. These responses were from native monolingual and non-Speakers of Spanish as well as Spanish-English bilinguals. The correlation between the two raters scores for the first prompt (*La familia que vivía en un faro*) was .877; the correlation between the two raters for the second prompt (*Susana la soñadora*) was .837.

Based on the unpublished version of the Spanish LAS Reading and Writing Technical Manual (1995), the authors report the following reliability correlation coefficients for the various subtests within Form 1A and 2A. The reliability coefficients for the various lectura and escritura subtests comprising Form 1A ranged between .80 and .90. The range for Form 2A was between .76 and .92.

Evidence for the construct validity (i.e., does the test measure what it purports to?) of the LAS-Oral Form 1B is based on three pieces of empirical evidence stemming from analyses of discriminate validity, predictive validity and test equivalency (De Avila & Duncan, 1982). Note that these analyses were conducted on the original Spanish version of the LAS developed in 1979 and not Form 1B published in 1990.

First, the analyses of discriminant validity demonstrated that the original Spanish version of the LAS could differentiate among fluent Spanish speaking, limited Spanish speaking, and non-Spanish speaking students based on Total Score and each of the five subtests making up the LAS-Oral.

The predictive validity analyses were actually conducted as part of a dissertation study. In this study a moderate positive correlation was found between the subject's score on the Crane Reading Test and their LAS-Oral Spanish proficiency score. Moderate positive correlations were also found between the subject's score on the Prueba de Lectura and the subject's LAS-Oral Spanish proficiency score.

The third piece of evidence supporting the construct validity of the LAS-Oral Form 1B stems from equivalency analyses between the LAS English and the original version of the LAS Spanish published in 1979. This study revealed that the distribution of scores was virtually the same for (151) English monolingual students and (102) Spanish monolingual students in third and fifth grade; however, a significant difference was found between the distribution of scores for first grade students. This is because the mean score for the Spanish monolingual speakers was significantly lower than the mean score for the English monolingual speaker.

Empirical evidence of the construct validity for the Lectura/Escritura LAS tests rests on three separate analyses. These data were described in the unpublished *Technical Manual* (De Avila and Duncan, 1995). First, the authors state that the range of correlations between the different parts of these tests ranged between .33 and .659. The same range encompassed analyses between parts of the test to the test as a whole.

The authors also conducted a series of analyses to examine the discriminant validity of the various subtests of Lectura/Escritura and of the tests as a whole. For Form 1A and 2A, students participating in these studies

were placed in either a Low (n=246) or High (n=308) Spanish literacy group. These analyses indicated that the High Spanish literacy group had a higher mean score than the Low group on every Lectura/Escritura subtest and the tests as a whole. However, whether the difference in the mean scores were statistically significant is not reported. For Form 2A, essentially the same results were found; the High group had a higher mean score on every subtest and test as a whole than the low group, but the authors did not report whether or not the differences in mean scores were significant.

Finally, the authors conducted a series of analyses which compared the scores between a "Native Educated" group and a "US Educated" group. As expected, the Native Educated students outperformed the US Educated students in every comparison. However, whether or not the differences in mean scores were significant is not reported in the yet unpublished Spanish LAS Reading and Writing Technical Manual (De Avila & Duncan, 1995).

Critiques

There do not appear to be any critiques of the Spanish version of the Language Assessment Scales tests described above. However, the reader may find the following critiques for English versions of the LAS of some use.

Carpenter, C.D. (1994). Review of the Language Assessment Scales, Reading and Writing. In *Supplement to Eleventh Mental Measurements Yearbook*. Buros Institute.

Guyette, T.W. (1994). Review of the Language Assessment Scales, Reading and Writing. In *Supplement to Eleventh Mental Measurements Yearbook*. Buros Institute.

Haber, L. (1985). Review of Language Assessment Scales. In J.V. Mitchell (Ed.), *The Ninth Mental Measurements Yearbook Volume I*. Lincoln, NE: University of Nebraska Press.

[\(table of contents\)](#)

Woodcock-Muñoz Language Survey

- [Purpose](#)
- [Age and language groups](#)
- [Administration time](#)
- [Cost](#)
- [Test administration](#)
- [Items](#)
- [Scoring](#)
- [Test design and theoretical foundation for test](#)
- [Reliability and validity](#)
- [Test critique references](#)

Purpose

As is the case in the *Comprehensive Manual for the Woodcock-Muñoz Language Survey*, this test will be

referred to as the "Language Survey" throughout this test handbook review. It is a recent arrival (1993) to the library of standardized language proficiency tests and a short acronym for the test has not been adopted yet by test users although the test authors sometimes refer to it as the LS-S for the Spanish version or LS-E for the English version. The review of this test is based upon the test materials and use of the test with a small selected group of examinees. At this writing, published test critiques and other sources of information about the Language Survey were not available.

Several purposes were identified for the Language Survey. It was specifically designed to measure CALP -- cognitive academic language proficiency. The Language Survey yields the following language proficiency classifications:

- Level 5: Advanced Spanish CALP
- Level 4: Fluent Spanish CALP
- Level 3: Limited Spanish CALP
- Level 2: Very Limited Spanish CALP
- Level 1: Negligible Spanish CALP

The *Comprehensive Manual for the Woodcock-Muñoz Language Survey* (referred to as the Manual hereafter) indicates that the wide range and breadth of test items allows the Language Survey to be used with confidence for the following purposes:

- To classify a subject's English or Spanish language proficiency,
- To determine eligibility for bilingual services,
- To help teachers understand a subject's language abilities,
- To assess a subject's progress or readiness for Spanish or English-only instruction,
- To provide information about program effectiveness, and/or
- To describe the language characteristics of subjects in research studies.

The *Manual* provides a paragraph description of each of these purposes and a rationale for how the Language Survey is able to meet each purpose.

Age and Language Groups

The Language Survey was designed to be administered individually to any student older than 48 months of age. It was not designed to be used with any particular language group and can be administered to students who speak any non-Spanish language to assess Spanish language proficiency. There is also an English version of the test that can be administered to students from any language group to measure their proficiency with the English language.

The norm group for the English form of the Language Survey included 6,359 subjects aged 24 months to 90 years of age. The *Manual* identifies and describes this norm group prior to a description of the process used to equate norms for the LS-S to the LS-E. Therefore, we will describe the norm group for the LS-E and then describe the equating process for the LS-S.

Subjects for the norm group were chosen randomly within a stratified sampling design. This design controlled for ten specific community and subject variables. The sampling design insured that the norm group subjects approximated the population distribution in the United States. Oversampling for small groups (for example, American Indians) was adjusted during data analyses to ensure that the norming data closely approximated the exact distribution in the U.S. population for all 10 norming variables. The ten variables used to construct the system of stratification were:

- census region,
- community size,
- sex,
- race,
- Hispanic,
- funding of college/university,
- type of college/university,
- education of adults,
- occupational status of adults, and
- occupation of adults.

In addition to the ten sampling variables used to stratify the normative sample, communities were selected with respect to four socio-economic status variables. These variables were:

- years of education,
- household income,
- labor force characteristics of adults, and
- occupation of adults.

The Language Survey norms are based on the distribution of scores at the subject's exact chronological age or grade placement resulting from a continuous-year norming procedure. Norms are based on data collected throughout the school year rather than on one or two data collection points during the year.

Norms for the LS-S are equated to the LS-E form of the Language Survey. The *Manual* states that, " This means that tasks underlying each LS-S test are scaled according to their empirical difficulty in English." (Woodcock & Muñoz-Sandoval, 1993, p.51). This is a 9 step process that involves the development of English items. These are Rasch-calibrated and normed. A similar bank of Spanish items is developed and Rasch-calibrated. The difficulty scale underlying the Spanish item bank was subsequently re-scaled, through an intermediate subset of parallel equating items, to the difficulty scale underlying the English item bank. The *Manual* identifies the following Spanish calibration sites:

Mendoza, Argentina

Madrid, Spain

San José, Costa Rica

San Ysidro, California

Chihuahua, Mexico

Miami, Florida

Guadalajara, Mexico

Mount Vernon, New York

Monterrey, Mexico

Houston, Texas

Lima, Peru

Pasadena, Texas

Ponce, Puerto Rico

Administration Time

The Language Survey requires approximately 15 to 20 minutes to administer the entire battery (four subtests). More time may be required with non-responsive young children. Some students may produce a scattering of correct responses requiring that a number of items be administered. Allow a reasonable amount of time for the student to respond and then move on to the next item even though the student may have responded correctly if given unlimited time to answer.

Cost

The initial investment in either form of the Language Survey is \$147 for the complete Spanish battery including manual, test book, 25 student test forms, and computer scoring and reporting program. If additional scoring forms are needed, the cost is \$21 for an extra package.

Test Administration

The Language Survey-Spanish is administered using an easel format. An easel format is a hard-bound book designed to fold open like a tent and sit on the table without being held. The examinee can see the test items on one side and the examiner can read the answer on the other side (which can not be seen by the examinee). The easel book allows the examiner to see both the items and the correct responses and allows for protecting the scoring form from the examinee's view behind the easel. Two or three administrations allow the examiner to become comfortable with the materials and administration of the test. It can be administered within 20 minutes using the "brisk" method of administration. However, students with a long wait time will take longer as will very young and very proficient students.

In most cases, the four subtests should be administered in the order they are presented in the test easel. However, the subtests can be administered in any order and a testing session can be stopped after the administration of any test. There are four subtests (referred to as "tests" in the Manual) and they are:

- Picture Vocabulary,
- Verbal Analogies,
- Letter-Word Identification, and
- Dictation.

Basal and ceiling rules are used to guide the amount of total time devoted to testing. Test items are arranged in order of difficulty with the easiest items first and the most difficult item last. Each subtest includes items that span a range of difficulty. By not administering items that are too easy or beyond the student's

capabilities, the number of items actually administered can be minimized. The basal and ceiling guidelines allow the examiner to estimate the score that would be obtained if every item were administered. Testing is conducted by complete pages in the easel book. As the examinee does not see items below the basal or above the ceiling, they are unaware that other test questions exist. This method of administration spares the examinee embarrassment when an item is too difficult.

Items

The Picture Vocabulary subtest presents the examinee with colorful drawings of things that might be considered typical and not-so-typical in the American school classroom, in homes, and in the mainstream, urban American community. The examiner asks the student to "Pon (ponga) tu (su) dedo en el gato". The gato would be one of several animals depicted. More difficult Picture Vocabulary items ask the examinee to provide the name for an object. Some of the more difficult items to be labeled are "monóculo" , "pagoda" , and "yugo".

The Verbal Analogies subtest presents printed analogies such as "mamá es a papá como hermana es a". The examinee must respond with "hermano" to complete this analogy correctly. Items become progressively more difficult and tap the examinee's knowledge of culturally specific analogies and the examinee's ability to reason through the use of analogy as well. The examiner prompts each item by reading, "Mamá es a papá como hermana es a (pause)." No other prompt is allowed.

The Letter-Word Identification subtest starts with simple drawings of several different items paired with one large realistic picture. The examinee is asked to point to the cartoon picture that tells about the big, realistic drawing. Later items on this subtest require the examinee to read letters and then words. The examinee is prompted with "Qué palabra es ésta? Sique (siga) con las otras. No vayas (vaya) demasiado rápido." The final subtest is Dictation. The examiner directs the student to write down responses to a range of written grammar and punctuation exercises on an answer sheet. Easy items require the student to copy a line or a circle after the examiner makes one. More difficult items require the examinee to spell words like "sábado" and abbreviations such as "del" for "de" y "el".

Scoring

The test booklets allow for quick collection of the test data and rapid scoring of answers. Item scoring is done during test administration. The manual provides clear directions on the process for transformation of the raw scores to scores for use in interpretation. Calculation of the raw score for each subtest is done after testing is completed. In all cases the test item is scored by placing a 1 or 0 in the appropriate space on the Test Record form; 1 is correct and 0 is incorrect.

Several scores are generated for each subtest, the two language clusters (oral, and reading-writing) and the overall test battery. They include a *W* score, age equivalents, grade equivalents, relative proficiency index, and CALP level scores. Explanations for how to convert raw scores to these other scores and definitions of these scores are included in the Manual. To those familiar with the Language Assessment Scales (LAS), the CALP level scores will look familiar. The *Manual* describes each kind of score in detail as well as how to interpret and apply the scores. The CALP level scores allow for quick program placement while the other scores provide more specific information about student strengths and weaknesses in particular areas (for example reading-writing). The Language Survey comes with computer disks that include a scoring program that will convert the raw scores for each subtest into *W* scores, grade equivalents and other types of scores. If

a computer is not available, the Manual provides specific and clear instructions for converting raw scores for the clusters to CALP levels or age and grade equivalents. The charts and tables needed for manual scoring are found in Appendices A and B in the *Manual*.

Test Design and Theoretical Foundation for the Test

The Language Survey uses Cummins' (1984) BICS and CALP distinction as the theoretical foundation for item selection and overall test design. BICS are basic interpersonal communication skills and CALP is cognitive academic language proficiency. The English and Spanish versions of the Language Survey are parallel forms of the same test and assess the students knowledge of "academic" language in either Spanish or English. The items included look like the type of language that would be used in classroom and advanced academic settings as opposed to the type of language used in everyday social exchange.

Reliability and Validity

There is extensive evidence in the manual for the reliability and validity of the English Language Survey. However, only one table (7-15, p. 70) presented separate information on the validity of the Spanish form of the Language Survey. Reliability information was not presented.

Evidence of validity was provided from a sample of 603 subjects from five grade levels attending school in Costa Rica, Mexico, Peru, Puerto Rico, and Spain. Total score on the Language Survey was correlated with achievement as measured by the *Batería Woodcock pisco-educativa en español* (Woodcock, 1982). The four broad achievement measures include Reading, Writing, Mathematics, and Total Achievement. A set of median correlations is provided in Table 7-15. The median correlation of the Oral Language Cluster was .71 with the Total Achievement score. The Broad Spanish Ability cluster had a median correlation of .83 with Total Achievement and had its lowest correlation with Mathematics (.65). In order to adequately assess the validity of any test, some measure of reliability must be established first as a test can not be valid if it is not reliable.

Critiques

As the Language Survey was released by Riverside Publishing Company in 1993, published test critiques are not available yet. The reader is urged to consider the types of issues raised in the test critiques published about some of the other tests reviewed in this handbook and apply the types of issues and questions raised to the Language Survey.

[\(table of contents\)](#)

Summary and Test Description Matrix

Summary

Five standardized Spanish language proficiency tests were reviewed. Four definitions of language proficiency were provided for the readers' consideration. In addition, the legal requirement to assess native (in this case, Spanish) language proficiency was described. The rationale for the review of the five tests selected (why these tests were chosen) was discussed. Finally, a section of the handbook provided the reader

with a brief description of each test and the test publisher's address and telephone number.

For each of the tests, the same selected topics were chosen for discussion. We chose to include pragmatic information related to the administration of the test, the time it takes to administer, test cost, test items, test scoring and the purpose for each test. Additionally, information was provided on the theoretical design of the test and the technical standards for the test. Two of the tests, the BINL and the BSM (I & II) assess only oral language (speaking). The other three (IPT, LAS, & the Woodcock-Muñoz Language Survey) assess reading, writing, listening and speaking. The information on each test depended on the amount of information provided in the manuals and test materials. Therefore, the descriptions of each test varied some related to the test documentation.

No judgement was made with regard to the "quality" of the test. These test reviews should **not** be interpreted as **test critiques**. We were unable to locate test critiques for the Spanish forms of these standardized tests. With the exception of the Woodcock-Muñoz Language Survey, all the English forms of the tests have been critiqued and a reference list of these critiques was provided at the end of each test section. The reader is urged to read the critiques about any specific test prior to purchase of the test, as well as, to consider the issues raised in the first chapter of this handbook with regard to Spanish language proficiency testing.

In reviewing the Spanish language proficiency tests, we found a substantial discrepancy between the amount of information available for these tests as compared to the English forms for the same tests. Not only were there no test critiques available for the Spanish forms, but there was also less evidence provided for the reliability and validity of these tests. Often, the test publishers offered information about the technical standards for the English forms without separate evidence that they had conducted reliability and validity tests for the Spanish forms. The reader needs to be aware that separate technical standards tests should be conducted and the results made available for the Spanish forms. They are different tests from the English forms even when they appear to "look" the same.

An additional problem we encountered in reviewing the Spanish tests concerned the construction of the Spanish norm group sample. Careful explanation for why the norm group was developed was not available for these tests. In fact, some of the tests did not detail the number of students included in the Spanish norm group. Again, we found that a description of the English norm group sample and the method for the development of the norm group was supplied to the reader in lieu of a description of the Spanish norm group. One of the tests supplied a list of countries from which the Spanish norm group sample was drawn but did not identify how many individuals from each country, number of males and females, ages, socio-economic status or other information that would allow the test consumer to identify the appropriateness of the test's norm group for the population to be tested.

It is important that the reader be an *informed consumer*. The testing industry and test products should be considered in much the same way you would consider purchasing any large and costly service or piece of equipment such as a car or a copy machine. In a report from the *National Commission on Testing and Public Policy* (1990), this statement is made, "Today, those who take and use many tests have less protection than those who buy a toy, toaster, or plane ticket" (p. 21). Carefully weigh the pros and cons before purchase and choose the product best suited to the needs of your students, teachers, programs, school and district.

Figure 1 includes contact information to access examination kits for any of the tests reviewed in this handbook. All tests reviewed have a toll free 800 number to call for ordering information. Usually, test

publishers allow a 30 day period free of charge in order to examine the test.

Test Descriptions and Publisher Information

Figure 1:
Five Standardized English Language Proficiency Tests Included in this Handbook

Assessment Instrument	General Description
<p>Basic Inventory of Natural Language (BINL)</p> <p>CHECpoint Systems, Inc.</p> <p>1520 North Waterman Ave.</p> <p>San Bernadino, CA 92404</p> <p>1-800-635-1235</p>	<p>The BINL (1979) is used to generate a measure of the K-12 student's oral language proficiency. The test must be administered individually and uses large photographs to elicit unstructured, spontaneous language samples from the student which must be tape-recorded for scoring purposes. The student's language sample is scored based on fluency, level of complexity and average sentence length. The test can be used for more than 32 different languages.</p>
<p>Bilingual Syntax Measure (BSM) I and II</p> <p>Psychological Corporation</p> <p>P.O. Box 839954</p> <p>San Antonio, TX 78283</p> <p>1-800-228-0752</p>	<p>The BSM I (1975) is designed to generate a measure of the K-2 student's oral language proficiency; BSM II (1978) is designed for grades 3 through 12. The oral language sample is elicited using cartoon drawings with specific questions asked by the examiner. The student's score is based on whether or not the student produces the desired grammatical structure in their responses. Both the BSM I & BSM II are available in Spanish and English.</p>
<p>Idea Proficiency Tests (IPT)</p> <p>Ballard & Tighe Publishers</p> <p>480 Atlas Street</p> <p>Brea, CA 92621</p> <p>1-800-321-4332</p>	<p>The various forms of the IPT (1978 & 1994) are designed to generate measures of oral proficiency and reading and writing ability for students in grades K through adult. The oral measure must be individually administered but the reading and writing tests can be administered in small groups. In general, the tests can be described as discrete-point, measuring content such as vocabulary, syntax, and reading for understanding. All forms of the IPT are available in Spanish and English.</p>
<p>Language Assessment Scales (LAS)</p> <p>CTB MacMillan McGraw-Hill</p>	<p>The various forms of the LAS (1978 & 1991) are designed to generate measures of oral proficiency and reading and writing ability for students in grades K through adult. The oral measure must be individually administered but the reading and writing tests can be administered in small groups. In general, the tests can be described as discrete-point and holistic, measuring content such as vocabulary,</p>

2500 Garden Road Monterey, CA 93940 1-800-538-9547	minimal pairs, listening comprehension and story retelling . All forms of the LAS are available in Spanish and English.
Woodcock-Muñoz Language Survey Riverside Publishing Co. 8420 Bryn Mawr Ave. Chicago, IL 60631 1-800-323-9540	The Language Survey (1993) is designed to generate measures of cognitive aspects of language proficiency for oral language as well as reading and writing for individuals 48 months and older. All parts of this test must be individually administered. The test is discrete-point in nature and measures content such as vocabulary, verbal analogies, and letter-word identification. The Language Survey is available in Spanish and English.

Note: Addresses and telephone numbers current as of November, 1995.

[\(table of contents \)](#)

References

- American Psychological Association. (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Amori, B. A. & Dalton, E.F. (1994). *IPT 1 Reading & Writing, Grades 2-3, Examiner's Manual, IDEA Reading & Writing Proficiency Test, Level 1, Spanish*. Brea, CA: Ballard & Tighe, Publishers.
- Amori, B. A. & Dalton, E.F. (1994). *IPT 2 Reading & Writing, Grades 4-6, Examiner's Manual, IDEA Reading & Writing Proficiency Test, Level 2, Spanish*. Brea, CA: Ballard & Tighe, Publishers.
- Anastasi, A. (1988). *Psychological Testing* (sixth edition). New York, NY: Macmillan Publishing Company.
- Bachman, L. (1990). *Fundamental Considerations in Language Testing*. NY: Oxford University Press.
- Basic Inventory of Natural Language. (1991). *Technical Report*. San Bernadino: CHECpoint Systems, Inc.
- Burt, M.K., Dulay, H.C., & Hernández Chávez, E. (1975). *Medida de Sintaxis Bilingüe. Manual Edición Española*. Harcourt Brace Jovanovich, Inc.
- Burt, M.K., Dulay, H.C., & Hernández Chávez, E. (1978). *Medida de Sintaxis Bilingüe II. Manual Edición Española*. Harcourt Brace Jovanovich, Inc.
- Burt, M.K., Dulay, H.C., Hernández Chávez, E. & Taleporos, E. (1980). *Bilingual Syntax Measure II Technical Handbook*. Harcourt Brace Jovanovich, Inc.
- Burt, M.K., Dulay, H.C., Hernández-Chávez, E., & Taleporos, E. (1980). *Bilingual Syntax Measure II*,

Technical Handbook. San Antonio, TX: Harcourt, Brace, Jovanovich, Inc.

Canale, M. (1984). On some theoretical frameworks for language proficiency. In C. Rivera (Ed.), *Language proficiency and academic achievement*. Avon, England: Multilingual Matters Ltd.

Canales, J. A. (1994). Linking Language Assessment to Classroom Practices. In R. Rodriguez, N. Ramos, & J. A. Ruiz-Escalante (Eds.) *Compendium of Readings in Bilingual Education: Issues and Practices*. Austin, TX: Texas Association for Bilingual Education.

Carrasquillo, A. (1988). Bilingual Syntax Measure II Test. In D.J. Keyser & R.C. Sweetland (eds.), *Test Critiques Volume VII*. Kansas City, MO: Test Corporation of America.

Collier, V. (1995). *Acquiring a Second Language for School*. Washington, D.C.: National Clearinghouse for Bilingual Education.

Council of Chief State School Officers (1992). *Recommendations for Improving the Assessment and Monitoring of Students with Limited English Proficiency*. Alexandria, VA: Council of Chief State School Officers, Weber Design.

CTB MacMillan McGraw-Hill (1991). *LAS Preview Materials: Because Every Child Deserves to Understand and Be Understood*. Monterey, CA: CTB MacMillan McGraw -Hill.

Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. In C. Rivera (Ed.), *Language proficiency and academic achievement*. Avon, England: Multilingual Matters Ltd.

Dalton, E.F. (1980, 1989). *IPT Examiner's Manual, IDEA Oral Language Proficiency Test, Spanish*. Brea, CA: Ballard & Tighe, Publishers.

Dalton, E.F. (1980). *IPT Technical Manual, IDEA Oral Language Proficiency Test, Spanish*. Brea, CA: Ballard & Tighe, Publishers.

Dalton, E.F. (1987). *IPT II Technical Manual, IDEA Oral Language Proficiency Test, Spanish*. Brea, CA: Ballard & Tighe, Publishers.

Dalton, E. F., Barrett, T.J., & Jensen, C.M. (1994). *IPT Reading & Writing Grades 2-6, Technical Manual, IDEA Reading and Writing Proficiency Test, Levels 1 and 2, Spanish*. Brea, CA: Ballard & Tighe, Publishers.

De Avila, E. & Duncan, S. (Forthcoming). Spanish LAS Reading and Writing. Technical Manual. CTB Macmillan McGraw Hill.

De Avila, E. & Duncan, S. (1990). Language Assessment Scales-Oral. Manual de Administración, Español, Forma 1B. CTB Macmillan McGraw Hill.

De Avila, E. & Duncan, S. (1990). Language Assessment Scales-Oral. Manual de Evaluación e Interpretación, Español, Nivel 1B. CTB Macmillan McGraw Hill.

Del Vecchio, A. & Guerrero, M. (1995). *Handbook of English Language Proficiency Tests*. Evaluation

Assistance Center-West. New Mexico Highlands University. Albuquerque, NM.

Duncan, S & De Avila, E. (1994). *Language Assessment Scales-Lectura/Escritura. Manual de Administración y Calificación, Forma 1A*. CTB Macmillan McGraw Hill.

Duncan, S & De Avila, E. (1994). *Language Assessment Scales-Lectura/Escritura. Manual de Administración y Calificación, Forma 2A*. CTB Macmillan McGraw Hill.

Durán, R.P. (1988). *Validity and Language Skills Assessment: Non-English Background Students*. In H. Wainer & H.I. Braun (Eds.), *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Figueroa, R.A. & García, E. (1994). *Issues in Testing Students from Culturally and Linguistically Diverse Backgrounds*. *Multicultural Education* (Fall), p. 10-18.

Hayes-Brown, Z. (1987). *Bilingual Syntax Measure II*. In J.C. Alderson, K.J., K.J. Krahnke & C.W. Stansfield (eds.), *Reviews of English Language Proficiency Tests*. TESOL.

Hernández Chávez, E. (1996). *The Four Skills Exam: Revisited*. Panel presentation at New Mexico Association for Bilingual Education Annual Conference. Albuquerque, NM.

National Clearinghouse for Bilingual Education. (1994). *Improving America's Schools Act. Title I: Helping disadvantaged children meet high standards*. George Washington University.

National Clearinghouse for Bilingual Education. (1994). *Improving America's Schools Act. Title VII: Bilingual education, language enhancement, and language acquisition programs*. George Washington University.

National Commission on Testing and Public Policy. (1990). *From Gatekeeper to Gateway: Transforming Testing in America*. Chestnut Hill, MA: National Commission on Testing and Public Policy.

Oller, J.W. Jr. & Damico, J.S. (1991). *Theoretical considerations in the assessment of LEP students*. In E. Hamayan & J.S. Damico (Eds.), *Limiting bias in the assessment of bilingual students*. Austin: Pro-ed publications.

Rivera, C. (1995). *How can we ensure equity in statewide assessment programs?* Unpublished document. Evaluation Assistance Center-East, George Washington University, Arlington, VA.

Roos, P. (1995). *Rights of limited English proficient students under Federal Law -- A guide for school administrators*. Unpublished paper presented at Weber State University, Success for all Students Conference, Ogden, UT.

Skutnabb-Kangas, T. (1988). *Multilingualism and the Education of Minority Children*. In T. Skutnabb-Kangas & J. Cummins (eds.), *Minority education: From shame to struggle*. Philadelphia: Multilingual Matters.

Solé, Y. (1990). *Bilingualism: stable or transitional? The Case of Spanish in the United States*. *International Journal of the Sociology of Language*, 84, pp. 35-80.

Spolsky, B. (1984). *The uses of language tests: An ethical envoi*. In C. Rivera (Ed.), *Placement procedures*

in bilingual education: Education and policy issues. Avon, England: Multilingual Matters Ltd.

Ulibarri, D., Spencer, M., & Rivas, G. (1981). Language proficiency and academic achievement: A study of language proficiency tests and their relationship to school ratings as predictors of academic achievement. *NABE Journal*, Vol. V, No. 3, Spring.

Valdés, G. and Figueroa, R. (1994). *Bilingualism and testing A special case of bias*. Norwood, NJ: Ablex Publishing Corporation.

Waggoner, D. (1994). Language Minority School Age Population now Totals 9.9 million. *NABE News*, 18, No. 1. National Association for Bilingual Education.

Waggoner, D. (1995). Are Current Home Speakers of non-English Languages Learning English? *NABE News*, 19, No. 3. National Association for Bilingual Education.

Wheeler, P. & Haertel, G.D. (1993). *Resource Handbook on Performance Assessment and Measurement: A Tool for Students, Practitioners, and Policymakers*. Berkeley, CA: The Owl Press.

Wong Fillmore, L. (1989). Language Learning in Social Context: The view from research in second language learning. In R. Dietrich & C.F. Graumann (eds.), *Language Processing in Social Context*. North Holland: Elsevier Science Publishers.

Woodcock, R. W. & Muñoz-Sandoval, A.F. (1993). *Woodcock-Muñoz Language Survey Comprehensive Manual*. Chicago, IL: Riverside Publishing Company.

[\(table of contents \)](#)

[Online Library](#)

[Subject Area: Assessment & Accountability](#)

The HTML version of this document was prepared by NCBE and posted to the web with the permission of the author/publisher.

[go to HOME PAGE](#)

www.ncela.gwu.edu