

Current Language Proficiency Tests and Their Implications for Preschool English Language Learners

Alberto Esquinca, David Yaden and Robert Rueda
University of Southern California

1. Introduction

Formal assessment of children of all ages is becoming increasingly common for a number of reasons. One of the major reasons is the demand for schools to be accountable. The “No Child Left Behind Act of 2001” includes accountability as one of its four main pillars, and stipulates that periodic testing is mechanism that will ensure that academic institutions are to show that children are making satisfactory progress. While some argue convincingly that the emphasis on accountability should be accompanied by a movement for greater equity in terms of the opportunities to learn to bring about significant and systemic changes (Gee, 2003), the drive for accountability is in place. For better or worse, the movement toward closing the achievement gap is using widespread testing as a means to ensure its goal, and so children will be tested more often and at a younger age.

Although preschool education has historically been a time in which children have an opportunity to socialize and prepare to transition into academic learning, the current emphasis is on beginning schooling as early as possible. As standardized testing of preschoolers becomes more common, it remains important and to consider the developmental needs of very young children. The National Research Council (2001:237) states that some of the developmental considerations that should be taken when testing very young children include young children’s inability to focus for extended periods of time, and their limited capacity to be purposeful and intentional. This means that in assessment situations, young children have a hard time attending to verbal instructions, situational cues and thus cannot be expected to complete test that school-aged children can.

Furthermore, the “No Child Left Behind Act” also emphasizes learning English by means of English-Only instruction to the exclusion of other methods, including bilingual education or primary language development and content skills (Rueda, in press). Thus English Language Learners will continue to be tested to ensure that they are making progress toward their linguistic goals. As Valdés (2001) notes, English language learning is often a priority over content learning.

Because of the great emphasis on assessment, we must recall that all assessment should strive to adhere to the principles of reliability and validity (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999). Test developers often strive to meet these goals, however, despite their best intentions, this goal is often not met. As Garcia and Pearson (1994) note, English language learners often face a mainstreaming bias in formal testing, including a norming, content, cultural and linguistic biases. Because of these know biases, Harris Stefanakis (1998:11) recommends that formal assessment of linguistic minority students be scrutinized for interpretation bias, construct validity of test translations, cultural bias in test items, and particular children’s test-taking experience.

The starting point for the present research is the need to understand the basic characteristics of language proficiency tests currently in use (the stated purpose of the tests, and other basic information), the reliability and validity information available, and the criticisms that those tests have received. While we are aware of the alternatives to standardized tests, we also know that, as August & Hakuta (1997) note, language proficiency tests cannot be ignored “because they are tapping an ability correlated with a variety of language proficiencies, not because they constitute language proficiency” (114).

The present paper constitutes the main findings from our search for the most appropriate assessments for English language learners. We will discuss some of the most common assessments used, overview the features of those tests, and discuss the most current and relevant issues that affect preschool English language learners.

2. Some commonly used instruments

According to the most recent *Survey of the States' Limited English Proficient Students and Available Educational Programs and Services 2000-01* (Kindler, 2002), school districts use a variety of methods to identify students as non-English proficient, place them in bilingual programs, and allow them to exit such programs (or reclassify them as English proficient). These methods include home language use surveys, criterion-referenced tests, achievement tests, and language proficiency tests. State Educational Agencies (SEAs) from the 50 states, the District of Columbia, Puerto Rico, Guam and the Northern Mariana Islands reported the use of language proficiency tests in each state. According to that ranking, the five most common language proficiency tests are the Language Assessment Scales (LAS), the IDEA Language Proficiency Tests (IPT), the Woodcock-Muñoz Language Survey, the Language Assessment Battery, and the Basic Inventory of Natural Language.

2.1 Common characteristics across instruments.

An examination of these measures reveals that the most used language proficiency tests are meant for school aged children. Because tests for prekindergarten children are not reported in the current survey our best guess as to what language proficiency tests are used with preschoolers are that (a) school districts continue to use assessments like the Pre-LAS, the Pre-IPT, and the Woodcock-Muñoz as the 1995 survey indicated (Macías, 1995) or (b) the above tests, meant for school-aged children, are inappropriately used with preschoolers.¹

Upon closer examination, we see that the four most-often used tests, the LAS, the IPT, the Woodcock-Muñoz, and the LAB share a number of features. The stated purpose for three of those tests is strikingly similar. With the exception of the Woodcock-Muñoz, the tests are designed to measure language skills (speaking, listening, reading, writing) and discrete aspects of language, i.e., lexicon, phonology, syntax. Table 2 gives details about some of the main characteristics of the four most commonly-used tests.

As Table 2 also shows, these language proficiency tests are part of a long tradition of language testing. Bachman's (2000) review of the literature on language tests outlines the development of language testing over the last 20 years. He points out that while testing practice from the mid-1960s and the 1970s tended to be based on a construction of language as skills (listening, speaking, reading, writing) and components (grammar, vocabulary, pronunciation), such constructions were critiqued as new approaches to the study of language emerged. Specifically, in the 1980s, the influence of communicative approaches to language instruction was paramount. Since applied linguists were developing approaches to teaching that focused on the co-construction of meaning, and the importance of context-based communication, traditional assessments (such as those developed in the 1970s) were ill-suited for the new approach. In the 1990s, test-makers became concerned with issues such as the development of (a) new research methodologies, such as criterion-referenced measurement, (b) practical advances, such as pragmatics testing, (c) factors that affect test performance, (d) authentic and performance assessment, and (e) ethical considerations of language testing.

¹ Test representatives do not have information or do not respond to inquiries about this kind of test misuse (personal communication).

Table 2: *Four most commonly-used language assessments*

Test	Purpose	Domains measured	Grades/ Ages	Items
LAS	“Designed to measure those English and language skills necessary for functioning in mainstream academic environments.”	LAS Oral: vocabulary, listening comprehension, story-retelling, LAS Rdg/Wrtng: mechanics and usage, vocabulary, fluency, reading comprehension, sentence completion, essay writing.	1-6	Vocabulary: “Name that picture” nouns (<i>desk, chair</i>); “Action words” (<i>eating, laughing</i>)
IPT	“To assist in the initial identification, designation, and redesignation of a student as being non-English Speaking, Limited English speaking or Fluent English Speaking.”	Vocabulary, comprehension, syntax, and verbal expression.	PreK-12	Vocabulary items include identification of nouns (<i>helicopter, stove</i>) and action verbs (<i>singing</i>).
WMLS	“Designed for measuring cognitive-academic language proficiencies.”	Verbal analogies, picture vocabulary, letter-word identification, dictation	4-adult	Vocabulary part tests over classroom-specific items. Verbal analogies consists of items such as: “ <i>sky... blue; tree...</i> ”.
LAB	"Designed to assess reading, writing, listening comprehension and speaking in English and Spanish."	Listening, speaking, reading, writing	Level 1: K-2	N/A

2.2 *Language constructs represented.*

The tests reviewed above are based on the assumption that language proficiency can be measured accurately by only sampling discrete aspects such as phonology, syntax, morphology, and lexicon. The tests rarely consider aspects of language that can be crucial to academic success, such as pragmatic competence (Cummins, 2000). In other words, most language proficiency tests limit the construction of language proficiency to grammatical competence.² An important flaw with this construction is that to assess grammatical competence, tests usually rely on prescriptivist notions of grammar. For instance, if one such type of test were to assess students’ acquisition of the English verb system, an item like (1) below might be presented.

(1) Dad called earlier. He _____ (*might/ is/ had/ might could*) stop by later this evening.

If a student were to fill in the blank with *might could*, he would probably be penalized because the Standard English verb system allows one modal verb in that position. However, if said student were a member of the group of native English speakers who make a distinction between (2) and (3) below, such an item would be invalid:

(2) He might stop by later this evening.

² The exception is, of course, the Woodcock-Muñoz, which is designed to assess Cognitive Academic Language Proficiency (CALP).

(3) He might could stop by later this evening.

While the differences in meaning are subtle and pragmatically determined, in (3) there is less likelihood that "he" will stop by than in (2) (Wolfram and Schilling-Estes, 1998:335). Speakers of the dialect in which sentences like (3) are common need contextual cues in order to distinguish the forcefulness of the assertion. However, a typical language proficiency test would not allow for nuances in meaning made by speakers of so-called non-Standard varieties of English.

Furthermore, to limit the construction of language proficiency to a closed set of grammatical categories negates the real need for language learners to master communicative principles which are essential in informal and academic contexts. After all, language learners must develop a range of communicative styles to suit their purposes. A language learner whose repertoire is limited to academic discourse styles cannot be considered fully communicatively competent.

Up to this point we have discussed how commonly-used tests utilize similar constructions of language proficiency, and how this construction of language proficiency is closely linked to prescriptivist notions of standard grammar. In the next section we discuss the criticisms that standardized language proficiency tests have received in test reviews.

2.3 Critiques of the four most commonly-used tests.

In addition to the limitation of language proficiency to grammatical competence, other criticisms are revealed in test reviews. These have indicated that some of the common shortcomings are (a) that many test items are not valid (Haber, 1985; Carpenter, 1994; Hedberg, 1995; Kao, 1998), (b) that interrater reliability is low (Crocker, 1998), and (c) that the tests are normed on populations that are not representative of the samples of children to whom these measures are commonly administered (Chesterfield, 1985; Haber, 1985; Shellenberger, 1985; Lopez, 2001). Table 3 includes a summary of the reviews.

Table 3: Critiques of four most commonly used tests

Test	View of language	Problematic aspects
LAS	Language consists of discrete skills and elements.	-Hedberg (1995): LAS-Oral is inadequate for placing language-minority students because of inadequate standardization procedures. -Carpenter (1994): LAS reading/ writing is inappropriate to make entry and exit decisions; teacher judgement would be just as valid.
IPT	Language consists of discrete skills and elements.	-Lopez (2001): Norming procedures limit test validity for a wide range of U.S. students, greater emphasis on discrete aspects of language proficiency and less emphasis on pragmatic competence, no studies were conducted to investigate how test content relates to achievement. -Ochoa (2001): Standardization sample is not representative of the range of U.S. English speakers, nor is the Spanish version representative of the range of Spanish speakers in the U.S..
WMLS	Cummins' BICS/ CALP distinction	-Crocker (1998): To account for construct validity, test makers rely on intercorrelations, not on an explanation of the underlying traits that test attempts to measure. -Kao (1998): Test-makers provide insufficient information about validity. There is little explanation about the Cognitive-Academic Skills (CALP) construct. -Schrank, Fletcher, and Guajardo Alvarado (1996):
LAB	Language consists of discrete skills and elements.	-Chesterfield (1985): LAS is problematic for identification of students for bilingual programs, contains unnecessary items, is inadequate to predict success or as a basis for intervention.

3. Implications and recommendations

3.1 Implications

In addition to the developmental recommendations made by the National Research Council noted above, it should be added that very young English language learners require assessments that are especially designed for them, in other words, that are developmentally appropriate and have been field-tested with representative populations (Meisels, 1994). For example, a test that is too long, or that requires preschool children to use paper and pencil is, for the most part, inappropriate. The National Research Council (2001) significantly notes that a young child's elicited utterances may not be representatives of their functioning, which is of particular concern when assessing language proficiency.

Although it should be clear that very young language learners have special needs, it is not clear why variations of the tests developed for school-aged children continue to be used (Pre-LAS, Pre-IPT). Although these tests can be given in a short time and easily scored, other assessment possibilities remain. McLaughlin, Blanchard, and Osani (1995) propose an assessment process based on a set of guidelines for preschool children. Because they stress that assessments should be developmentally and culturally appropriate, their assessment process takes into consideration the child's linguistic background (such as sociolinguistic variety), and is constructed to allow children to demonstrate what they can do. In short, their assessment process is an alternative to standardized testing toward authentic, instructionally embedded instruments. Information is collected about children's meaningful language use in categories such as: use of language to express feelings, use of language to express ideas, use of language to ask for help, use of language to describe, among others. The information is collected in a portfolio, a narrative summary about the child's linguistic development is written. Finally, the information is conveyed to interested agents, and used to develop the curriculum.

McLaughlin, Blanchard and Osani's process approach is significantly different from the standardized testing approach represented by the tests critiqued above. Not only is assessment conceptualized as process, instead of as product, but also the functions that assessment fulfills are different. In the assessment-as-process approach information is easily usable by teachers, students, parents and other agents. Assessment here fulfills a *formative* function. In the assessment-as-product approach information is very useful for administrators and fulfills a *summative* function.

In stark contrast to the kind of information that can be gathered from formative assessments, there is the case of a standardized language proficiency test given to preschoolers, the Pre-LAS *español*, which we have seen that provides information about a narrow range of linguistic constructs. In a review of the construct validity of the *Pre-LAS español*, MacSwan, Rolstad, and Glass (2002) argue that the parts of the test that constitute most of the variance in the scores and which are responsible for the non-Spanish speaker rating do not, in fact, relate to oral language proficiency in Spanish. Because the test puts students into categories that range from fluent Spanish speaker to limited Spanish speaker and non-Spanish speaker, students who spoke non-standard varieties of Spanish or who were unresponsive were categorized as non-Spanish speaker. The authors use convincing statistical evidence to show that two thirds of those students identified as non-Spanish speakers were incorrectly identified.

The above two examples of an authentic language assessment process and a standardized language proficiency test show a revealing contrast. While information gathered from standardized testing gives information about a limited range of linguistic aspects, and it can be deceptively simple to interpret, it is often used by in contexts that extend beyond the classroom, for instance for political gain or in media campaigns.³ In contrast, authentic assessments provide information about a wider range of language uses, should be easily interpretable, and is effectively used in the classroom.

³ The *L.A. Times* reported that thousands of children in Los Angeles Unified School District were not proficient in either English, nor their home language.

3.2 Recommendations.

It is strongly recommended that any a single measure (often administered on a single occasion) not be used as the sole determiner for screening, reclassification and program exit (transition) decisions. These types of decisions must be based upon triangulation of instruments—in a portfolio-type fashion. While we agree that any one test of language ability may “potentially” offer valuable information for educators, a complete measure of language proficiency can only be made with a network of instruments and observations of language use in a variety of situations over time. Furthermore, having a series of indicators of language development ensures that a test is not invalidated by a young child’s unfamiliarity with testing situations, his or her lack of control over his linguistic resources, or a child’s language non-Standard variety (dialect).

In addition, test-makers, developers and consumers must continue to develop an awareness of the appropriate and ethical use of language tests. While the drive has traditionally been to develop tests which can withstand evaluations of validity, test-makers must continue to be concerned with the value that standardized tests are given, a trend attributed to Messick (1989:20), who justly raises the issue of the social implications of the application of test results. He extended the notion of validity of interpretations of measures to include how test results are used, thereby making test-makers accountable for the tests they construct. Thus, in the case of young English language learners, it can justly be said that standardized language proficiency tests should be designed and used with extreme caution. Furthermore, the measured of proficiency that the tests give should be useful for all interested parties, including parents and children.

4. Conclusion

The purpose of the present paper is to examine the features of the most common standardized language proficiency tests using test reviews which analyze issues such as construct validity, and reliability. We believe that in the current context, in which there is not only a greater use of standardized tests, but also greater value placed on them, it is imperative to be critical of the measures that are used to assess students. The major findings of this paper are that the most often used standardized tests construct language proficiency as “skills and components,” a view which is out of tune with the latest findings in the field of sociolinguistics. Finally, we strongly recommend that the use of these tests be limited to one of a constellation of measures of proficiency.

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards of educational and psychological testing*. Washington, D.C.: American Psychological Association.
- August, D., & Hakuta, K. (1997). (Eds). *Improving schooling for language-minority children: A research agenda*. Washington, D. C.: National Academy Press.
- Bachman, L. (2000). Modern language testing at the end of the century: Assuring that what we count counts. *Language testing* 17 (1) 1-42.
- Bowman, B. T., Donovan, M. S., & Burns, M. (Eds.). (2001). *Eager to learn: Educating our preschoolers*. Washington, D.C.: National Academy Press.
- Burt, M.K., Dulay, H.C., Hernández-Chávez, E., and Taleporos, E. (1980). *Bilingual Syntax Measure II, Technical handbook*. San Antonio, TX: Harcourt, Brace, Jovanovich.
- Carpenter, C. D. (1994). Review of *Language Assessment Scales, Reading and Writing. Supplement to the eleventh mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Chesterfield, K. B. (1985). Review of *Language Assessment Battery. The Ninth Mental Measurements Yearbook Volume I*. Lincoln, NE: University of Nebraska Press.
- Crocker, L. (1998). Review of the *Woodcock-Muñoz Language Survey. The Thirteenth Mental Measurements Yearbook*. Lincoln, NB: University of Nebraska Press.

- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Clevedon, UK: Multilingual Matters Ltd.
- Cummins, J., Muñoz-Sandoval, A.F., Alvarado, C.G., & M.L. Ruef (1998). *The Bilingual Verbal Ability Tests*. Itasca, IL: Riverside.
- Dalton, E. F. (1991). *IPT Oral Grades K-6 Technical Manual, IDEA Oral Language Proficiency Test Forms C and D English*. Brea, CA: Ballard & Tighe, Publishers.
- De Avila, E.A. & Duncan, S. E. (1990). *Language Assessment Scales, Oral Technical Report, English, Forms 1C, 1D, 2C, 2D, Spanish, Forms 1B, 2B*. Monterey, CA: CTB MacMillan McGraw-Hill.
- Del Vecchio, A., & Guerrero, M. (1995). *Handbook of language proficiency tests*. Albuquerque, NM: Evaluation Assistance Center–Western Region, New Mexico Highlands University.
- Garcia, E. (1985). Review of *Bilingual Syntax Measure II. The Ninth Mental Measurements Yearbook Volume I*. Lincoln, NE: University of Nebraska Press.
- Garcia, G.E. and Pearson, P.D. (1994). Assessment and diversity. *Review of research in education* 20:337-391.
- Gee, J.P. (2003). Opportunity to learn: A language-based perspective on assessment. *Assessment in education* 10:27-46.
- Guyette, T. (1985). Review of Basic Inventory of Natural Language. *The ninth mental measurements yearbook Volume I*. Lincoln, NE: University of Nebraska Press.
- Guyette, T. (1994). Review of Language Assessment Scales, Reading and Writing. *Supplement to the eleventh mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Harris Stefanakis, E. (1998). Whose judgement counts?: Assessing bilingual children, K-3. Portsmouth, NH: Heinemann.
- Haber, L. (1985). Review of Language Assessment Scales. *The ninth mental measurements yearbook Volume I*. Lincoln, NE: University of Nebraska Press.
- Hedberg, N. L. (1995). Review of Language Assessment Scales-- Oral. *The twelfth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Kao, C. (1998). Review of the *Woodcock-Muñoz Language Survey. The Thirteenth Mental Measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Kindler, A. (2002). *Survey of states' limited English proficiency students and available educational programs and services: 2000-2001 summary report*. Washington, D.C.: National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs.
- Lopez, E. A. (2001). Review of the *IDEA Oral Language Proficiency Test. The Fourteenth Mental Measurements Yearbook*. Lincoln, NE: University of Nebraska Press.
- MacSwan, J. Rolstad, K. and Glass, G.V. (2002). Do some school-age children have no language? Some problems of construct validity in the Pre-LAS Español. *Bilingual research journal* 26: 213-238.
- Macías, R. (1998). *Summary Report of the Survey of the States' Limited English Proficient Students and Available Educational Programs and Services 1995-96*. Washington, D.C.: National Clearinghouse for Bilingual Education.
- McLaughlin, B., Gesi Blanchard, A., & Osanai, Y. (1995). *Assessing language development in bilingual preschool children*. Washington, D.C.: National Clearinghouse for Bilingual Education.
- Messick, S. (1988). Validity. In R.L. Linn (Ed.) *Educational measurements. Third edition*. New York: American Council on Education/ McMillan.
- No Child Left Behind Act*. (2001). Retrieved October 2, 2002 from <http://www.nochildleftbehind.gov>
- Ochoa, S. H. (2001). Review of the *IDEA Oral Language Proficiency Test. The Fourteenth Mental Measurements Yearbook*. Lincoln, NE: University of Nebraska Press.
- Rueda, R. (in press). Student learning and assessment: Setting an agenda. In Pedraza, P. and Rivera, M. (Eds.). *National Latino/a Education Research Agenda Project*.
- Shellenberger, S. (1985). Review of *Bilingual Syntax Measure II. The Ninth Mental Measurements Yearbook Volume I*. Lincoln, NE: University of Nebraska Press.
- Tidwell, P.S. (1995). Review of Language Assessment Scales-- Oral. *The twelfth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Valdés, G. and Figueroa, R. A. (1994). *Bilingualism and testing: A special case of bias*. Norwood, NJ: Ablex.
- Valdés, G. (2001). *Learning and not learning English: Latino students in American schools*. New York: Teacher's College Press.
- Wolfram, W., & Schilling-Estes, N. (1998). *American English*. Malden, MA: Blackwell.
- Woodcock, R. W. and Muñoz-Sandoval, A.F. (1993). *Woodcock-Muñoz Language Survey Comprehensive Manual*. Chicago: Riverside Publishing Company.

ISB4: Proceedings of the 4th International Symposium on Bilingualism

edited by James Cohen, Kara T. McAlister,
Kellie Rolstad, and Jeff MacSwan

Cascadilla Press Somerville, MA 2005

Copyright information

ISB4: Proceedings of the 4th International Symposium on Bilingualism
© 2005 Cascadilla Press, Somerville, MA. All rights reserved

ISBN 978-1-57473-210-8 CD-ROM
ISBN 978-1-57473-107-1 library binding (5-volume set)

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Press.

Ordering information

To order a copy of the proceedings, contact:

Cascadilla Press
P.O. Box 440355
Somerville, MA 02144, USA

phone: 1-617-776-2370
fax: 1-617-776-2271
sales@cascadilla.com
www.cascadilla.com

Web access and citation information

This paper is available from www.cascadilla.com/isb4.html and is identical
to the version published by Cascadilla Press on CD-ROM and in library binding.