DOCUMENT RESUME

ED 363 090 FL 021 315

AUTHOR Canales, JoAnn
TITLE Assessment of Language Proficiency: Informing Policy and Practice.
INSTITUTION Southwest Educational Development Lab., Austin, Tex.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 89
CONTRACT 400-86-0008
NOTE 29p.; Appended materials may not reproduce well.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; *Educational Policy; Elementary Secondary Education; *Language Proficiency; *Language Tests; *Policy Formation; *Second Languages; Standardized Tests; Statewide Planning; Student Evaluation; Testing; Verbal Tests

ABSTRACT

A discussion of language proficiency focuses on the conceptual framework for assessing proficiency and its implications for educational policy formation at the state and local levels. First, the concept of language is defined in terms of the interaction of these elements: language subsystems, communication skills, language domains, language registers, and knowledge of the language. Context of language use is also considered. Two common language assessment approaches, oral language proficiency tests and standardized achievement tests, are then examined for advantages and limitations, and an alternative, integrative approach is outlined. The latter approach is based on the sociolinguistic perspective that language is more than the sum of its discrete parts, and uses measures of integrative skills, including rating scales, interviews, dictation tests, and cloze tests. Finally, guidelines for policy and practice at the elementary and secondary school levels are proposed, targeted specifically at state policy-makers, the state education agency, and the local education agency. Appended materials consist of notes on the characteristics of selected tests of English language proficiency and results of a study of the reliability of a number of language proficiency tests. (Contains 11 references.) (MSE)

ASSESSMENT OF LANGUAGE PROFICIENCY:

INFORMING POLICY AND PRACTICE

JoAnn Canales, Ph.D.
North Texas State University

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement, U.S. Department of Education, under Contract Number 400-86-0008. The contents of this publication do not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

ASSESSMENT OF LANGUAGE PROFICIENCY: INFORMING POLICY AND PRACTICE

JoAnn Canales

SUMMARY

This paper seeks 1) to establish a common conceptual framework for language proficiency assessment by addressing five fundamental questions about the process; 2) to describe an "integrative approach" for conducting language assessment and its role in the educational process; and 3) to discuss the implications of suggested guidelines for making informed decisions about policy and practice for state-level policymakers, for decisionmakers in state education agencies and for administrators and practitioners in local education agencies.

Language proficiency assessment – how to assess, when to assess – has been a long-standing dilemma for both theoreticians and practitioners. Addressing this complex issue requires a common conceptual framework defined by such questions as:

- What is language?
- In what context is language assessment needed?
- What is involved in assessing language proficiency?
- What do existing oral language proficiency tests tell us about a student's ability to perform in the classroom?
- What do standardized written tests tell us about a student's ability to perform in the classroom?

All educational service providers, from policymakers to practitioners, must understand these fundamental questions if they intend to use language assessment practices as instruments that inform decision-making rather than as biased descriptions of students' educational opportunities and successes.

Conceptual Framework

Defining "Language"

Language is a multi-dimensional concept that requires the interaction of linguistic subsystems, communication skills, domains of language, registers, and knowledge of the language.

Linguistic subsystems refer to the four subsystems that include graphophonemics (letters/sounds), lexicon (vocabulary), morphology (grammar and word order), and semantics (meaning). These subsystems exists in all languages. Levels of proficiency may differ between any or a combination of these subsystems. For example, an

individual's pronunciation of English may be difficult to understand and word order may not be grammatically correct, but, the person may, nonetheless, be capable of verbally expressing a complex concept. Thus, the individual's language proficiency may be "high" with respect to vocabulary and meaning but "low" regarding the use of sounds and grammar/syntax.

"...a non-native speaker of English ... might be able to understand a set of oral directions, but would have a difficult time comprehending the same directions if they were written."

Communication skills are used in pairs that can be categorized as either receptive (listening, reading) vs. expressive (speaking, writing) or oral (listening, speaking) vs. print (reading, writing), as illustrated below. In either case, the pairing of the skills does not suggest that proficiency in one skill (e.g., listening) necessarily equals proficiency in its paired skill (i.e., speaking or reading). For instance a non-native speaker of English whose language learning efforts have not focused on oral production (speaking) or on academic reading materials easily might be able to understand a set of oral directions, but would have a difficult time comprehending the same directions if they were written.

CATEGORIES OF COMMUNICATION SKILLS

	Oral	Print
Receptive	Listening	Reading
Expressive	Speaking	Writing

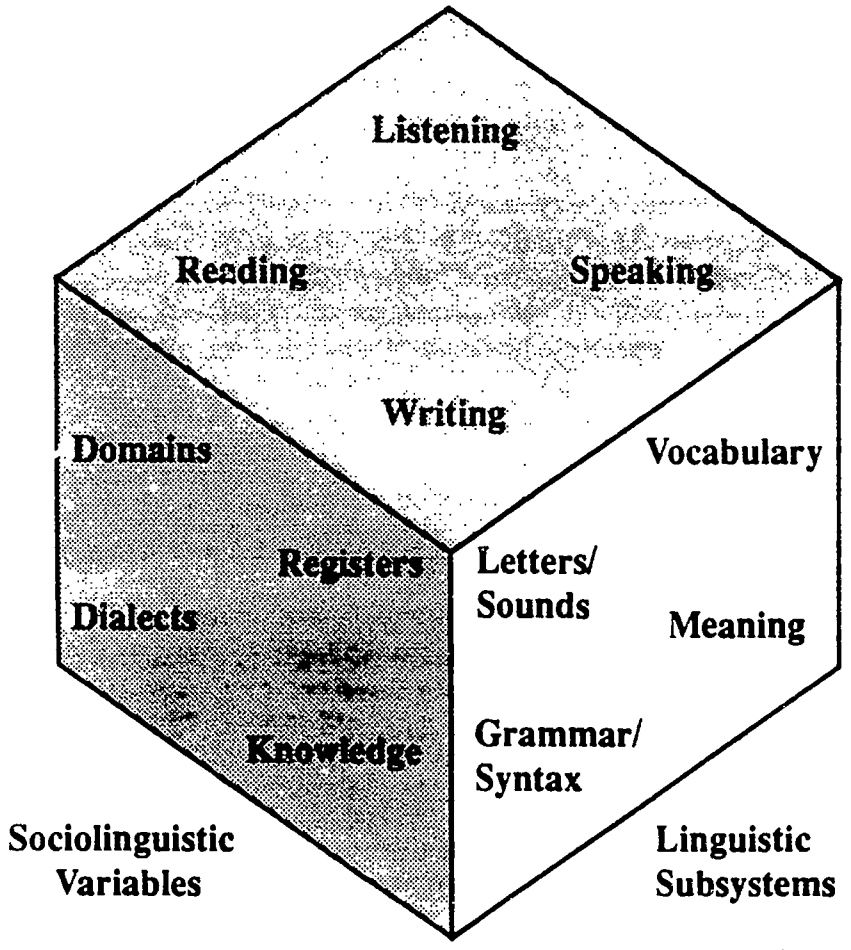
Domains of language include four categories: home, peers, school, and community. Understandably, this aspect of language suggests that the language spoken within the context of the four domains may differ, not only in the pronunciation of the words, but also in the vocabulary used and the level of complexity required of the utterances (phrases/sentences). The content of discourse (connected speech) and the vocabulary used must vary from setting to setting if comprehension is the purpose of the communication.

Registers of a language refer to a range of expressive opportunities that depend on such factors as situations, participants, relationships, roles, topics, and locale (Fishman, 1972). For example, the formality of a teacher's conversation differs considerably, depending on whether he or she is talking with students, colleagues, or the school principal.

Dialects refer to the forms of speech employed by a given community. This may consist of different words, or shortened, or altered forms of words, found in the native language such as "mon" for "man" commonly used by natives of the Caribbean of Bahamian Islands.

Knowledge of the language refers to the extent an individual has experience in a given language. The average monolingual child from a family where there are frequent opportunities for verbal exchange, including family discussions and reading of children's literature, enters kindergarten at age five with at least 20,000 hours (365 days x 12 average hours of wakefulness x 5 years) of exposure to both the receptive and expressive skills of the language spoken at home, with peers, or in the community.

Communication Skills



"Communicative competence, for any given situation, is a function of what we have learned about language, its structure, and its vocabulary for that particular situation."

There are numerous possible interactive combinations between the various components of a language, as illustrated below. Most likely some individuals will be more proficient in some of these possible combinations in a given language, and less proficient in others.

Here's an example of the situation. Two people, one a car owner and the second a mechanic, are looking at an automobile that "won't run." After studying the situation under the hood for several minutes, both are asked to explain the problem. No matter how familiar the owner is with the car, the mechanic, because of training, is conversant about engines and automobiles at a level of detail far exceeding that of most car owners. As a result, the mechanic's explanation of the problem will be far more articulate and linguistically competent than that of the car owner.

Similarly, a student, who may demonstrate some knowledge of a language on the playground, cannot be assumed to be communicatively competent in that language in the classroom without specific instruction (Cummins, 1980). Communicative competence, for any given situation, is a function of what we have learned about language, its structure, and its vocabulary for *that* particular situation.

Given the complexity of language and the various opportunities in which language may be used, it is important to understand the context in which language will be used before determining what constitutes proficiency. Appropriate determination of an individual's language knowledge base is a tri-fold process that includes:

1. Identifying the context in which a student's language proficiency should be determined. This context will provide the vocabulary and the level of complexity that students must work with if they are to be successful in the calssroom.

2. Determining a student's level of knowledge in the four linguistic subsystems. In terms of instruction, this information will help teachers identify the subsystems and the level at which a student must be taught.

3. Assessing a student's level of proficiency in the four communication skills to be utilized in the classroom. This will allow the teacher to structure classroom activities appropriately.

Language Assessment Approaches

Language assessment practices now common in schools include the use of an oral language proficiency test (OLPT) to assess skills in listening and speaking[1] and a standardized achievement test to assess skills in reading and writing. The OLPT is used primarily to place and exit students in pre-kindergarten through first grade; both OLPTs and standardized achievement tests are used in grades 2 through 12. Reviews of school district practices show that, for the most part, these measures are used in a summative manner (administered annually or at the beginning and end of the student's tenure in a program) and are seldom used to monitor student progress on a formative (on-going) basis.

How appropriate are each of these measures as assessments of language proficiency?

The OLPTs typically are **discrete-point measures** of language because they assess a person's knowledge of a language's structures using a single test item, such as subject-verb agreement or forming plurals. Instruments used as OLPTs have been reviewed in recent years by many agencies and experts (Texas 1988, 1985, 1979, 1977; California, 1982, Rivera and Simich, 1981; Ulibarri, et al, 1981; Thonis, 1980). The primary purpose for reviewing these instruments has been to determine how appropriate and efficient OLPTs are with students whose first language is not English. Among the findings were:

- The data reported for purposes of determining validity and/or reliability was insufficient and, at times, non-existent.
- The language characteristics of children on whom the test was normed were not identified.
- The number of discrete skills covered was minimal. (Of the more than 200 dimensions of language that might be appraised, most currently available language instruments assess fewer than ten.)

[1] In recent years, some language proficiency measures have been developed that assess reading and writing in addition to listening and speaking, e.g., the Language Assessment Battery (LAB), the Language Assessment Scales (LAS), and the Maculaitis Assessment Program (MAC). While they do measure many more skills than the oral language proficiency measures, the training, administration time, and scoring time involved have precluded their extensive use in the schools.

"(M)any (standardized) tests claim to determine an individual's proficiency in English after a 15-to-20-minute test – a dubious claim given the complexity of the language and the many contexts in which it is used."

- The scaling of scores to determine proficiency was relatively arbitrary and overlapping.

- None of the OLPTs addresses student's knowledge of academic language (the language to be used in the classroom).

Many of these instruments require a minimal amount of training to administer them. Many are accompanied by claims that use of the test can be self-taught, and they make little mention of any unique characteristics the examiner should have in order to obtain valid and reliable results. Thus, unlike ability tests such as the Wechsler Intelligence Scale for Children, Revised (WISC-R), which must be administered by certified examiners, OLPTs can be given by virtually anyone – aides, parents, older children – and frequently are, regardless of the administrator's formal training.

Moreover, many of the these tests claim to determine an individual's proficiency in English after a 15-to-20-minute test – a dubious claim given the complexity of the language and the many contexts in which it is used.

Furthermore, these tests are not comparable language assessment instruments (Ulibarri, 1981). The basis for the Bilingual Syntax Measure (BSM) is knowledge of syntax (word order): examinees are shown pictures and asked questions, and their responses are recorded. The Language Assessment Scales (LAS) relies heavily on a person's ability to retell a story presented by the examiner or a tape. The Basic Inventory of Natural Language (BINL) involves showing visual stimuli and having individuals tell what they see. Students' scores on these three tests vary significantly and, as a result, are unreliable in classifying language proficiency. Differences between the various instruments have implications for entry into programs for children whose native language is not English, since they yield different scores depending on the different discrete points they measure.

In addition to OLPTs, a student's scores on the reading and language arts subtests of **standardized achievement tests** often are used to assess reading and writing proficiency in English. These tests are also considered discrete-point measures since their construction is such that they present a single stimulus – a test question – on a predetermined academic skill to be measured, such as identifying the main idea.

On the surface, the use of these instruments seems logical since one

obviously must have a certain level of proficiency in a language in order to perform well on an achievement test in that language. Unfortunately, the nature of standardized testing generally precludes such tests from being valid measures of a student's proficiency in a language.

First and foremost, standardized tests are not designed to be instruments of language proficiency assessment; they are designed to rank order students on the basis of their knowledge of academic skills. Second, few, if any, limited English proficient students are likely to be included in the norming of the test; thus, scores are *not* representative of the population on whom the test is being used as a test of language proficiency. Third, several kinds of students are not likely to do well on standardized measures of academic achievement for various reasons. Students likely to score below minimum competence (that is, below average) include students who:

- come from low-income families,
- have reading disabilities,
- have emotional disabilities,
- are linguistically different,
- lack basic skills, and/or
- have poor test-taking skills.

"If there is little or no relationship between curricula and test objectives, then there will be little or no relationship between a student's academic proficiency and the test scores."

Because a number of reasons can contribute to low scores, using a standardized achievement instrument as a measure for determining language proficiency can result in the improper placement of a student in a program.

A related caveat: if standardized achievement scores are to be truly valid, they must reflect achievement on content covered in the classroom. If there is little or no relationship between curricula and test objectives, then there will be little or no relationship between a student's academic proficiency and the test scores. This is an extremely important consideration when using standardized achievement tests, regardless of their language, as measures for identifying, placing, and exiting students from special programs.

The widespread use of these two types of instruments, OLPTs and standardized achievement tests, as criteria for determining placement and exit, can be attributed to two probable factors:

1. To date there are no truly standardized measures of oral language proficiency and

2. OLPTs and standardized tests offer the appearance of quick, "objective" screening tools for making educational decisions for placement.

The major fault, however, is that all of such instruments that have been reviewed have been found minimally acceptable for use with limited English proficient students and thus, should **never** be the sole source(s) for determining a child's educational placement. When OLPTs and achievement tests must be used as a standard for measuring language proficiency, additional measures incorporating systematically obtained teacher judgment and observation data should accompany the "standard information."

An Integrative Approach to Language Assessment

The limitations previously discussed are not sufficient to totally discard the concept of language assessment. The theoretical underpinnings of language assessment instruments need to be revisited, however, and they must be couched in the more realistic disciplines of sociology **and** linguistics (sociolinguistics), in addition to the historical disciplines of psychology **and** linguistics (psycholinguistics). The sociolinguistic perspective of language acknowledges the fact that language usage:

"...language is more than just the sum of its discrete parts."

- is dynamic and contextually based (varies depending upon the situation, the speakers, and the topic)
- is discursive (requires connected speech)
- requires the use of integrative skills to achieve communicative competence.

The sociolinguistic theoretical viewpoint suggests that language is more than just the sum of its discrete parts, and therefore, a similar philosophy for language assessment instruments. Instruments consistent with this philosophy are at the opposite end of the continuum from discrete-point measures and are known as *measures of integrative skills*. They include rating scales, interviews, dictation tests, and cloze tests.

Rating Scales

Classroom observations of students interacting in various settings are the basis for determining students' linguistic proficiency. A student's linguistic performance in listening **and** speaking is rated on

a five-point scale of proficiency, ranging from non-native speaker of English to proficient speaker of English, for each of the four linguistic subsystems (graphophonemic, lexicon, morphology, and semantics). These rating scales are completed by the classroom teacher after observing students in various settings. Separate rating scales can also be completed for observations of casual, social interactions, such as playground or cafeteria talk, compared to observations of formal academic interactions, such as reading or social studies lessons. Appropriate completion of these rating scales requires that the classroom teacher have an understanding of the criteria used to rate each of the linguistic subsystems.

Interviews

Structured interviews are developed and administered on an individual basis. Ideally, an examiner should conduct the interview while a language specialist transcribes the examinee's responses, noting the use of the four linguistic subsystems. The advantages of this kind of measure are that it can be individually tailored to the experiences of the examinee and it allows the examiner opportunities to explore an individual's knowledge of the language.

The disadvantages, however, are several. First, it usually requires two people to administer the interview, a skilled interviewer and a language specialist. Second, this interview scenario has the potential to distract the examinee and perhaps contribute to diminished responses because of intimidation, especially for young children. Third, individualized administration makes it a time-consuming procedure. Finally, without appropriate scaling criteria, interviews are unsuitable for widespread use in schools as a tool for identification and placement of students.

Dictation Tests

The examinee listens to text dictated from graded material and writes down what is heard. The premise for this measure of integrative skills is that the individual needs to have knowledge of the four linguistic subsystems in order to convert speech to print. The advantages of dictation tests include:

- They are easily developed from material used in everyday classroom situations such as basal readers, science books, or social studies books.
- They can be administered in a group setting.
- They do not require extensive specialized training to develop or administer.

The few disadvantages of dictation tests, which can occur in the administration phase and the scoring phase, are manageable if the examiner is aware of them. First, an examiner's dialectal differences may cause the examinee difficulties in transcribing speech to print, a problem that could be overcome by using a taped version of the dictation. A related problem, students' lack of familiarity with this type of test, can be mitigated with practice sessions prior to the actual dictation to be used as the measure of language proficiency.

Second, an examinee's unfamiliarity with all of the variations in spelling of English sounds may cause interference for the examinee in converting speech to print – writing "miss is esmith" for "Mrs. Smith," for example. This difficulty can be overcome by having the dictation tests scored by someone who knows the differences between the graphic and phonetic systems of the examinee's native language compared to the systems in English.

Third, the dictation test requires that the individual being tested knows how to write.

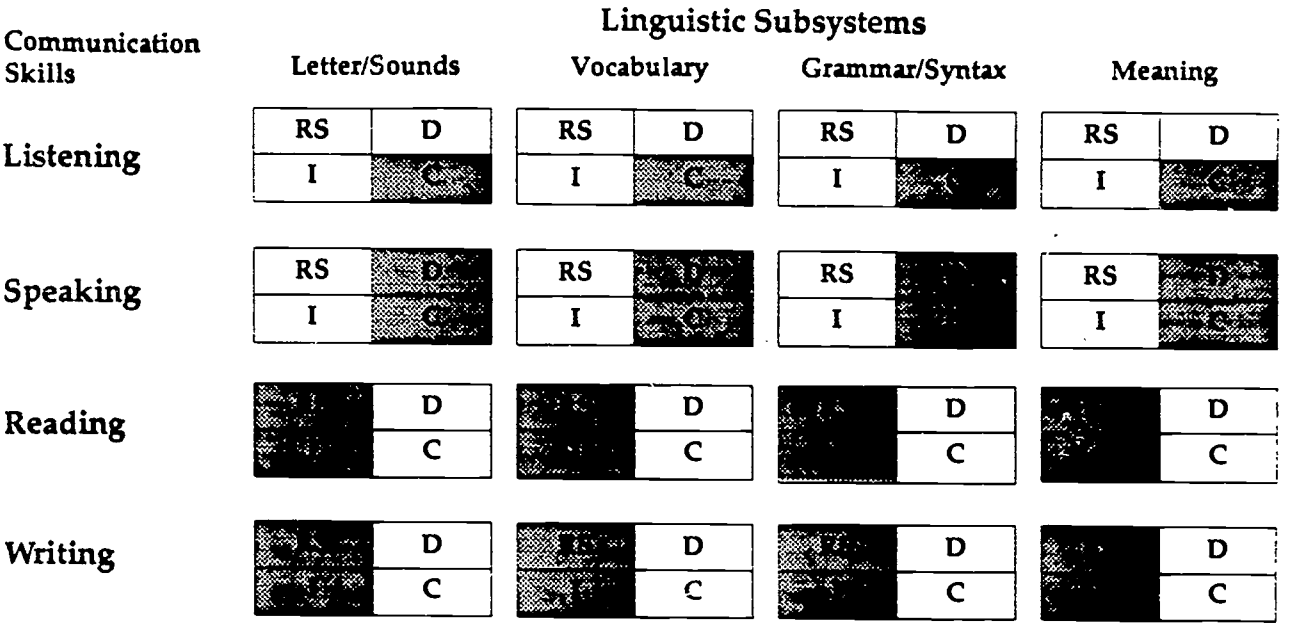
Finally, appropriate criteria for scaling need to be developed, as in the case of the interviews.

Cloze Tests

The examinee is asked to complete a readability-graded passage from which words have been omitted at regular intervals (usually every fifth word). The premise of this procedure is that language is highly redundant, with many contextual clues that can inform the examinee of the appropriate missing words if that person has a command of the language being tested. Cloze tests have been used for many years and validated by reading specialists. Administered and analyzed properly, the results of cloze tests will yield information regarding the examinee's level of facility with the text. Such information is useful in planning for students' instructional needs.

In addition to its instructional orientation, there are many advantages to this procedure. The test can be prepared easily using texts that students use in the classroom, thus making the assessment procedure a functional one. Further, the test can be administered in a group setting and quickly scored. If administered to native speakers at the same grade level, their scores can serve as a basis of comparison for the non-native speakers' scores. Additionally, the construction, administration, and scoring of the cloze test do not require any extensive specialized training to use correctly.

MEASURES OF INTEGRATIVE SKILLS



Non-shaded measures are appropriate for the skills and substructures indicated. The following codes are used for the measures:

RS = Rating Scales, D = Dictation, I = Interview, and C = Cloze.

All four of these measures assess more than one discrete, isolated language or linguistic skill as shown in the following chart.

The selection of an appropriate instrument for use in a school setting requires considerations other than the language compenents such as administrative costs, training required, and ease of adminstration and scoring. One way to use these factors in selecting among the methodologies is to rate the measures on each of the criteria, using a three-point scale that represents relative costs in dollars or effort (i.e., 1 = least cost, 2 = medium cost, 3 = most cost). Ratings for all criteria are then tallied for each measure, permitting a numerical comparison of total cost considerations.

A typical set of such pragmatic ratings for the various measures of integrative skills is presented below. These ratings are realistic for many school districts, and they suggest that rating scales and cloze tests will be less time-consuming and less expensive than dictation or interviews. Such ratings may vary, however, depending on a district's available resources.

With cost-consideration rankings in hand, district administrators can better assess the cost-effectiveness of employing alternative measures for specific purposes, weighing both costs and benefits.

PRAGMATIC RATINGS OF MEASURES OF INTEGRATIVE SKILLS*

Measure	Cost Factors (dollars and/or effort) Development	Duplication	Training	Testing	Scoring	Relative Cost
Rating Scales	1	1	2	1	1	6
Dictations	2	2	1	1	3	9
Interviews	2	1	3	3	3	12
Cloze Tests	1	2	1	1	1	6

*Rating may differ based on resources available in district.

Guidelines for Policy and Practice

As the primary decisionmakers regrading the schooling of children, educators are faced with many opportunities either to enhance or preclude a child's successful experiences in school. These opportunities are multiplied in the case of linguistically different children who require bilingual/English as a Second Language instruction. This field of education is fraught with primitive assessment tools, unclear definitions for effective instructional practices, and an insufficient number of trained specialists to meet growing needs. Following are some suggestions for service providers at the various levels of policymaking – legislative, state education agency (SEA), and local education
agency (LEA) – to begin the process of overcoming these seemingly insurmountable obstacles.

State Policymakers

Historical precedent has shown that, without legislative intervention, effective school practices may not always be equitable. To ensure equity in the schooling process, legislators must take an active role in sponsoring legislation that enables the creation of programs that are flexible, sensitive to student needs, and above all, funded. Legislators need to consider:

1. Mandate appropriate indentification and placement of students whose first language is not that of the school's.

2. Fund the identification and instructional processes necessary to meet the needs of language-minority students.

3. Establish consequences for state and local education agencies (SEA/LEA) that do not appropriately address the needs of language-minority students.

State Education Agency

The state education agency can address the issue of language assessment in a variety of ways, including the following:

1. Set parameters for identifying, placing, and exiting students on the basis of English language proficiency.

2. Organize a statewide task force composed of a cross-section of educators – administrators, speech therapists, bilingual/ESL teachers, and regular teachers – to assist with setting parameters and developing assessment criteria and/or instruments.

3. Develop or modify a language-proficiency rating scale for use statewide to ensure consistent criteria for determining listening and speaking proficiency.

4. Develop training materials for use in local school districts to ensure that all local personnel understand the use of and can administer the rating scale properly

5. Develop criteria for cloze procedures to ensure that all LEAs develop cloze tests using the same readability measure and following the same principles of development (e.g., omit or include *a* or *an*; include or exclude technical terms in content-area cloze tests).

6. Pilot-test language-assessment procedures to ensure their comprehensibility, as well as their feasibility, in the field.

7. Monitor language-assessment practices at the LEA level periodically to ensure appropriate use of criteria and/or instruments.

8. Review language-assessment practices at the LEA level periodically to determine need for modification of criteria and/or instruments.

Local Education Agency

Since the implementation of language-assessment procedures rests with local education agency (LEA) staff, LEAs should:

1. Provide training for all administrators and teachers in the district.

2. Administer the assessments consistently and periodically at pre-determined times.

3. Monitor implementation of language-assessment procedures.

4. Review outcomes of student performance in language proficiency assessment committees.

5. Provide constructive feedback to state education agency staff regarding language assessment practices.

Some optional considerations include:

1. Participate in the SEA's state-wide task force to develop criteria/instruments for language-proficiency assessment.

2. Participate in pilot testing language assessment practices.

These suggestions are congruent with those for state education agency staff. In addition, local agency staff have additional responsibilities since ultimate decisions regarding educational intervention rests with them. These responsibilities require that LEA staff weigh all available information prior to making determinations about a child's educational placement, particularly when exiting the child from a program. In the case of the bilingual child, a number of factors must be considered (Mace-Matluck, 1988):

1. Does the oral language proficiency test used by your school district measure not only the kind of language needed in your class but also that which will be needed at the next higher grade level?

2. Has the student sufficiently mastered the basic language skills that will prepare him or her to deal successfully with the shifting emphasis of language skills at the next level of schooling?

3. Are you familiar with the textbooks that the student will be expected to use the next year? Are you sure he or she can handle both the language and content demands of these books with a minimum of help?

4. Have you exposed the student to a wide variety of vocabulary, text types, and topics?

5. What is the student's reading rate in English? How well does the student comprehend both content-area and reading texts?

6. What are the student's scores in language arts and reading on the most recently administered standardized achievement test? Are the scores at least as high as the average student in that student's school (i.e., do the scores compare favorably with the school or district norm?)

7. Have you looked at the student's scores on the math, science, and social studies sub-tests of the standardized achievement test? Do they compare favorably with the school or district norm?

8. How high is the student's anxiety level in your class? Is school stressful, or is the student self-confident and able to handle frustration or failure?

"Because language assessment is predicated on a dynamic, context-based tool, namely language, it too must be dynamic and context-based."

These are basic questions that must be addressed prior to making decisions to exit students from a bilingual/ESL program.

One final word: Language assessment, like language itself, should not be a static, cast-in-concrete practice. Because language assessment is predicated on a dynamic, context-based tool, namely language, it too must be dynamic and context-based. Otherwise, assessment and subsequent interpretations will be unrealistic, isolated observations of language proficiency, which may be totally unrepresentative of the student's language skills.

REFERENCES

Cummins, Jim. "The Entry and Exit Fallacy in Bilingual Education." **NABE Journal**, Volume III, 1980, pp. 25-29.

Erickson, Joan G. and Donald R. Omark. **Communication Assessment of the Bilingual Bicultural Child**. Baltimore: University Park Press, 1981.

Fishman, J.A. "The Sociology of Language" in **Language and Social Context**, ed. P.P. Giglioli. Great Britain: Penguin Books, 1972.

Hernandez-Chavez, Eduardo, Marina Burt, and Heidi Dulay. "Language Dominance and Proficiency Testing: Some General Considerations." **NABE Journal**, Volume III, No. 1, Fall, 1978.

Mace-Matluck, Betty. "Consequences of Exiting Students Too Early fom Special Language Assistance Programs." Austin, Texas: Southwest Educational Development Laboratory, 1988.

Oller, John W. "Discrete-Point Tests Versus Tests of Integrative Skills" in **Focus on the Learner: Pragmatic Perspectives for the Language Teacher**. Rowley, Massachusetts: Newbury House Publishers, Inc. 1973.

Report of the Committee for the Evaluation of Language Assessment. Austin, Texas: Texas Education Agency, 1977, 1979, 1983, 1985, 1988.

Rivera, Charlene, ed. **Placement Procedures in Bilingual Education and Policy Issues**. Clevedon, England: Multilingual Matters LTD., 1984.

Rivera, Charlene and Carmen Simich. "Issues in the Assessment of Language Proficiency of Language Minority Students." **NABE Journal**, Volume VI, No. 1, Fall, 1981.

Thonis, Eleanor. "Language Assessment: A Malignant Good." Speech delivered in Honolulu, Hawaii, July 24, 1980.

Ulibarri, D. M., M.L. Spencer, and G.A. Rivas. "Language Proficiency Tests and Their Relationship to School Ratings as Predictors of Academic Achievement." **NABE Journal**, Volume III, Spring, 1981.

Appendix A

Characteristics of Selected Tests of English Language Proficiency

Instruments identified as appropriate by U.S. Office of Education

Prepared by ECIA Chapter 1 Evaluation Technical Assistance Center
Revised by SDE 1988

TEST	DESCRIPTION	FORMS, LEVELS GRADES	PUBLISHERS/ DISTRIBUTOR
Language Assessment Scales (LAS) 1977-82-88	Measures student performance across four linguistic subsystems. An audio-cassette is used for 4 of 5 subtests to ensure standard pronunciation. Examiner records students responses.	Forms: English A & B Spanish Pre LAS: Kindergarten & 1st Level I: 2-5 Level II: 6-12 LAS Written (English) LAS Reading (English)	Published by: CTB/McGraw-Hill Karen Woods - Evaluation Consultant 2619 East Orange Tempe, Arizona 85281 (602) 833-8599
Basic Inventory of Natural Language (BINL) 1977-79	Oral language test of natural speech patterns. Student's verbal response is recorded and transcribed for computer analysis.	Forms: two for each level Kit A or B: K-6 Kit C or D: 7-12	Distributed by: CTB/McGraw-Hill Karen Woods - Evaluation Consultant 2619 East Orange Tempe, Arizona 85281 (602) 833-8599
Bilingual Syntax Measure (BSM) 1975-78	Picture tests include 25/26 items. Examiner asks questions about pictures, scores student's answers, and enters score in student response booklet.	Forms: English, Spanish Level I: K-2 Level II: 3-12	The Psychological Corporation Larry Burnell - Area Representative (303) 690-3025 555 Academic Court San Antonio, Texas 78204-2498 1 (800) 228-0752
IDEA Oral Language Proficiency Tests (IPT) 1979	Tests include 81/91 items organized by levels of proficiency. Student responds orally to verbal and picture stimuli until level of proficiency is reached.	Forms: English (A & B) Spanish Portugese Level I: K-6 Level II: 7-12	Ballard & Tighe 480 Altas Street Brea, CA 92621 (714) 990-4332 (800) 321-4332 Judy Shaffstall - Area Representative

Other Language Proficiency Instruments must be identified, described and submitted for approval by SDE

Appendix B

In July, 1982, the Assessment of Oral English Proficiency: A Status Report was presented at the Fifth International Symposium on Educational Testing at the University of Stirling in Scotland. Authors of the report are Dr. David Ramirez, Evaluation and Research Consultant, California State Department of Education; Dr. Barbara Merino, Assistant Professor of Education and Linguistics at the University of California; Dr. Thomas Bye, Director of Bilingual Education at the Vallejo Unified School District, California; and Norman Gold, Ed.D. Bilingual Education Consultant, California State Department of Education.

The paper was based on the work of the Language Proficiency Instrument Review Committee for the California State Department of Education whose work centered around three major tasks.

1. To adopt standards based on linguistic and psychometric principles by which the technical properties of oral English language proficiency tests designed to identify children of limited English proficiency.

2. To apply these standards to the selection and recommendation of instruments for purposes of this identification process.

3. To develop specific recommendations for improving language proficiency tests.

The committee prescreened many oral English language tests using six specific criteria. Next, they thoroughly analyzed those instruments which met the criteria along a common set of characteristics which were:

I. Test Background
II. Qualitative Evaluation of Test Materials
III. Administration
IV. Validity
V. Reliability
VI. Normative Standards
VII. Interpretation.

Recommendation by the California State Department of Education for an individual test "was based largely on information provided in support of its validity, reliability, normative standards, and interpretation. That is, respectively, how well were the following questions addressed.

1. Does this test adequately assess oral English language skills?

2. Does it do so consistently?

3. Does this test accurately assess the oral English language skills needed by language minority children to function successfully in California's English-only classrooms?

4. Does this test accurately discriminate between non-, limited-, and fluent-English-speaking children?" (p.3 of California report.)

The findings of this report are included as information for those commonly used instruments which assess English oral production for Hispanic, Native American, and Asian students in Title VII projects in the state. Please refer to this report for more information.

CRITERION VALIDITY CLAIMS OF ORAL LANGUAGE PROFICIENCY TESTS

	Concurrent Validity	Predictive Validity
BSM I & II	No cleims. Independent evidence of moderate relationship with other oral measures. Virtually no correlation with academic achievement.	No claims. No evidence.
LAS I & II	Moderate correlation with achievement. Moderate to fairly high correlations with other oral measures and teacher judgement.	Predictive velidity claims are really concurrent. No evidsnce.
IPT	Moderate to high correlations with other oral measures and teacher judgements.	No claims or evidsnce.
BINL	Mixed correlations reported with ESL continuum and separate comprehension test of unknown characteristics and small n's. Independent evidence of lower correlation with other oral meesures than shown among those other measures. Low and negativs correlations with achievement.	Low to moderate correlations between BINL at pretest and ESL mastery test 10 months later. Neither n's nor characteristics of ths ESL test specified.

TEST-RETEST RELIABILITY CLAIMS OF ORAL LANGUAGE PROFICIENCY TESTS

No information provided

LAS II — Missing

Information given but not adequate or appropriate as noted

BINL — Different sets of pictures used in two administrations. (Specific sets not known.) The effect of pictures and occasion are confounded.

BSM I — Low coefficient (KAPPA .44; 66 percent agreement) suggest instability of scores.

Information given is satisfactory but not complete as noted

BSM II — For grades 4-8 only, students (n = 85) mainly in upper levels of proficiency (76 percent were FES) Pearson correlation of .82 to .84.

LAS I — Correlation coefficients low to moderate .36-.64, though on a small sample (n = 29) and different scorers and examiners were used at both test administrations.

Information thorough and adequate

IPT — Pearson r= .94, n= 218. Age, sex, grade level, locale, and teacher opinion of academic ability are provided. Change scores also included.

INTER-SCORER RELIABILITY CLAIMS OF ORAL LANGUAGE PROFICIENCY TESTS

No information provided

BSM II — Missing

Information given is satisfactory but not thorough

BSM I — Inter-judge agreement values of .73 are reported. Written protocols were used.

BINL — Correlation between machine vs. human scores on 162 sentences is reported. It is not clear how many different children produced sentences. Human raters included a linguist and speech therapist (highly skilled raters of the sort not likely to be found in the field.) The machine functioned under what appears to be two different programs since some discrepancies were noted in two machine scores.

LAS I — Inter-rater correlations between teachers, aides, and school psychologist. Ethnicity, age, and number of children in sample are noted, but language proficiency level is not. Amount or quality of training received by scorers are not specified. Reported difficulties experienced in the field in scoring storytelling subtest are not addressed. Correlations ranged from .85 to .96.

LAS II — Inter-rater reliability involved two linguists trained by LAS personnel (an unusual situation), not likely to be found in the field. However, correlations by rater not highly trained was respectable (r= .71). In story

retelling, other subtests had coefficients in the .90's.

IPT — Different examiners as well as different raters were used. Correlations are high (sample of n= 218).

INTERNAL CONSISTENCY RELIABILITY CLAIMS OF ORAL LANGUAGE PROFICIENT TESTS

No information provided

BSM I — Missing
BINL — Missing

Information is satisfactory but not complete

IPT — Correlation coefficients are high (Cronbach Alpha of .98 for test as a whole) point biserials are also reported for each item. No data provided for subsca' .s.

LAS I — Data available on monolingual-English speakers and children from different ethnic backgrounds. Coefficients ranged from a low of .39 for vocabulary subtest to a high of .95 for comprehension. The relationship of different stories used in story retelling is not discussed nor is the relationship of story retelling to other subtests.

LAS II — Coefficients ranged from .65 on vocabulary subtest to .93 on phonemes subtest. Data drawn from a monolingual Anglo sample. As in LAS I, story retelling has not been analyzed.

BSM II — High internal consistency on a sample of 500. Separate coefficients computed for level 3/4 and level 5/6 items. Point biserials for individual items are not provided.

New Mexico Department of Education
Recommended procedures for exemption from state testing program
[illegible] 6.9.1.3

STEPS	PROCESS	RESULTS
STEP 1: IDENTIFICATION OF HOME LANGUAGE	- Home language questionnaire - Teacher language observation - Student language survey	LANGUAGE OTHER THAN ENGLISH
STEP 2: ASSESSMENT OF ORAL ENGLISH PROFICIENCY	- Select instrument - Train teachers in use of instrument - Administer test - Record scores	LAS LEVEL 1,2 IPT LEVEL A,B BSM LEVEL 1,2 BINL LEVEL 1,2
STEP 3: PROCEDURES FOR EXEMPTION FROM STATE TESTING PROGRAM	- Document scores - Request parent approval - Specify alternative assessment	REQUEST EXEMPTIONS FROM STATE SUPERINTENDENT
STEP 4: REFINEMENT OF CURRICULUM	- Base curriculum design in test results - Include ESL/ELD, native language, and culture to teach competencies - Continue to assess progress, diagnose needs, modify program	SUBMIT SUMMARY REPORT TO STATE SUPERINTENDENT

28