TRENDS IN ORAL PROFICILNCY TESTING

Roseann Dueñaz Gonzalez
Center for English as a Second Language
The University of Arizona
Tucson, Arizona

True to the Saussurean definition of language as a "system of systems," the traditional testing of all areas of language has been devoted to the atomistic view of assessment in which the recognition and production of the components of each language skill have been isolated by test constructors. Just as the atomistic approach to foreign language teaching has shifted to a more holistic approach, so has the emphasis in testing moved from the possibly more objective atomistic discrete-point approach to the more subjective holistic integrative approach (Carroll, in Allen, 1965). The emphasis is now on what Dell Hymes labels "communicative competence" (in Gumperz and Hymes, 1970), what Rivers (1972, p. 73) defines as the "ability to communicate, to interact verbally which presumes some knowledge (cognition) both in the perception of units, categories, and fuctions, and in the internalizing of the rules relating to these categories and functions." Most language tests, however, limit themselves to the measurement of these isolated components and do not concern themselves with the contention that communicative competence (the ability to handle language as a means of communication) is more than the sum of its linguistic components (Hymes, in Gumperz and Hymes, 1970).

General Language Testing

Lado's (1961) work Language Testing was the culmination of the testing procedures that were initiated, formalized, and developed during approximately 20 years of the audiolingual tradition of foreign language teaching and testing. In his work, Lado discusses specific teachniques for testing the "elements of language" namely pronunciation, intonation, stress, grammatical structure, and vocabulary. His work, dedicated to Charles Fries, incorporates Fries's dedication to the hierarchial framework of language. Fries (1953, p. 26) emphasized the "recognition and production" of "the features of arrangement" that

constitute the structure and sound system of the language; therefore, the assessment of language proficiency consisted of the deductive and analytic isolation of the recognition and production of each of the "features of arrangement."

In a summary statement about language testing made in 1961, John B. Carroll (in Allen, 1965, p. 365) observes that "the work of Lado and other language testing specialists has correctly pointed to the desirability of testing for very specific items of language knowledge and skill judiciously sampled from the usually enormous pool of possible items." He suggested a model (Figure 3) for the isolated abilities that may be measured (Carroll, in Allen, 1965, p. 367).

	Language Aspect			
Skill	Phonology or orthography	Morphology	Syntax	Lexicon
Auditory comprehension				
Oral production				
Reading				
Writing				

Figure 3. Model for Testing Isolated Abilities

Oral Proficiency Testing

The testing of oral proficiency has traditionally dealt with the testing of a system of separate and discrete functions; oral production of sound segments, recognition of sound segments, oral production of intonation, stress, and juncture patters, and the recognition of the same; oral production of grammatical elements, and the recognition of same. Representative of the attention paid to the testing of isolated elements such as suprasegmentals are the two separate chapters dedicated to stress and intonation in Language Testing (Lado, 1961).

The variety of techniques Lado (1961) suggested to test oral production may be classified into two types: The first, direct; and the second, indirect:

1. The use of highly structured speech samples (live or mechanically produced), rated according to very specific criteria.
2. The use of paper and pencil objective tests of pronunciation presumably providing indirect evidence of speaking ability.

The use of highly structured speech samples in the testing of oral production in the estimation of many has relegated oral proficiency to the area of pronunciation. Yet the oral reading of a written language sample by the student was referred to by Lado (1961, p. 83) as "the most uniform, precise, and simple method for testing production of sound segments of a language." Harris (1969) suggests the following as an examplary test of a highly structured speech sample:

1. The subject must read aloud a word, a sentence, a paragraph, or a prepared speech.
2. The oral reading is either recorded for delayed scoring or it is simultaneously scored.
3. The scorer listens to the pronunciation of a set number of elements per sentence, marking the incorrect pronunciation of the two variables.
4. A second part of the exam may consist of the subject reading a passage aloud, while the scorer listens for stress, voicing, vowel quality, series intonation, intonation contours, consonant clusters, and pitch.
5. The third section of the exam may consist of sentence conversion, wherein the subject is asked to convert sentences from positive to negative, statements, to interrogatives, from past to present, etc. On the basis of the grammatical acceptability of each utterance, the test is scored.

Rand (1963) incorporates the structured speech sample technique into his "A Short Test of Oral English Proficiency." He varies the format by adding an initial section consisting of open-ended questions such as "Where have you taught school and where do you now teach?" That section is followed by a section in which the subject was asked to convert an affirmative statement to a negative. Rand (1963, pp. 208-209) listed

the following as the grading criteria actually given the student:

(1) Utilization of time: (a) Do you only answer the questions or do you elaborate on the answer? (b) Do you utilize all the time allowed?

(2) Variety of sentence patterns: Do you use complex and compound sentences, only simple sentences, or incomplete sentences?

(3) Correctness of pronunciation (especially intonation, pitch, stress, and rhythm).

(4) Correctness of grammar.

The conversion section was scored according to the following scores from 5, as the highest, down to 1 as the lowest, with an additional score of x for no attempt to answer, as follows:

5 Native intonation, and speed. Little trace of foreign accent.

4 Fluent, but not native intonation. Some segmental difficulty.

3 Word-by-word. (too many primary stresses).

2 Word-by-word. Much segmental difficulty. Pitch, stress, and junctures bad.

1 Unintelligible.

x No attempt (Rand, 1963, p. 209).

Another method of evaluating "speaking ability" suggested by Lado and supported by many is the mimicry method," credited by Adelaida Paterno (in Allen, 1965, p. 383) to be the "simplest and most dependable of all oral proficiency tests." A statement is made by the test giver and the subject repeats the statement, whereupon the test giver scores the utterance. A variation of this method is the mechanical recording of the session for scoring at a latter time; however, the .31 correlation of most mimicry tests has rendered them, at the least suspect (Jones and Spolsky, 1975, p. 27).

Current attempts to improve mimicry tests include Buiten and Lane's (1965) Speech Auto-Instructional Device, capable of extracting pitch, loudness, and rhythm parameters from short spoken phrases and comparing these to internally-stored criteria of accuracy. Pulliam (1969) has described the development of an experimental speech interpreter, also computer-based, which can evaluate the examinee's pronunciation of specific short utterances. The drawbacks to the use of such devices include equipment cost and complexity.

An indirect method to test speaking proficiency is the paper-and-pencil test based on the assumption that (1) "what the foreign learner 'hears' himself say silently is, in fact, what he says aloud, and (2) that a sufficiently broad range of pronunciation problems can be tested by this indirect method to allow us to generalize about a subject's overall control of the English sound system" (Harris, 1969, p. 88). An often-used technique in the paper-and-pencil test is the use of rhyme words, word stress and phrase stress. In the rhyme word test, the examinee is presented with a test word; and then, from a list of possible rhyme words, he must select either a rhyming or nonrhyming word. In the word and phrase section, the examinee is required to identify the correct stress on a word or phrase by choosing the answer.

The validity of such tests is questionable. The tests are reported to be inconsistent with the examinee's actual competence in the language as evaluated by more direct means. There is a problem with the basic technique in that a student may be choosing the correct response by pronouncing the stimulus and rhyme word incorrectly. Another criticism is that the total sound system of the language is inadequately represented.

In 1969, ten years after the publication of Lado's Language Testing, Harris (1969, p. 85) observed that in the area of oral production "highly structured speech sample testing showed "much promise," and that the "validity of paper-and-pencil objective techniques remains largely unproven." Decidedly contrary to Lado, Harris (1969, p. 90) calls the "scored interview though not as reliable a measure as we would wish for, . . . the best technique for use in relatively informal, small-scale testing situations." Harris' Testing English as a Second Language reflected only a miniscule movement toward less isolation of variables and a greater emphasis on the totality of oral production testing. His greatest contribution to the field of oral production testing is

the devotion of almost an entire chapter on the oral interview. That serious attention to the integrative approach of the oral interview marks the beginning of another era in the field of oral proficiency testing.

The Integrative Approach to Oral Proficiency Testing

Oral proficiency testing from approximately 1969 to the present has been characterized by the tendency toward what may be called a more integrative, holistic approach. The following are the factors influencing this trend:

1. A concern with communicative competence.
2. A concern with integrative versus discrete-point test items.
3. A concern with functionality.
4. A concern with the notion of overall proficiency.

A discussion of each of these factors follows, after which experimental and working tests which reflect any one or any combination of these factors will be discussed.

Communicative Competence

In his article entitled "Are We Really Measuring Proficiency with Our Foreign Language Tests?" Briere (1971) expressed a concern for the inadequacy of present testing methods. He called for continued work to develop instruments capable of measuring "communicative competence," as opposed to measuring only "linguistic competence." The concept of "communicative competence," and extension of the Chomskyan sense of competence, includes the capacity of the native speaker to produce or understand utterances which are appropriate to the context in which they are made (Hymes, in Gumperz and Hymes, 1970). "Learning a language includes not only understanding and using the grammar of that language; it also includes the capacity to use the language in a way that is appropriate to the situational and verbal constraints operating at any given time" (White, 1974a, p. 130).

Paulston (1974), Jakobovits (1970), and Rivers (1972) distinguished between "the narrow notions of linguistic competence" and the broader " sociolinguistic and psycholinguistic" implications of communicative competence. Jakobovits (1970) suggests that communicative competence includes three levels of linguistic phenomena: the linguistic level, the implicit level, and the implicative level. The latter two differentiate communicative competence from lingustic competence. By "implicit meaning," he refers to the eliptically derived conceptual event which an utterance represents. By "implicative meaning" Jakobovits refers to the information in an utterance about the speaker himself, his intention, his psychological state, his definition of the interaction.

Jakobovits (1970) suggests three ways of testing communicative competence according to his scheme:

1. Judgments of acceptability: the subject judges the acceptability of an utterance or picks the most appropriate of two similar utterances.
2. Semantic differential techniques: the subject rates a word on a seven-point bipolar adjectival scale.
3. Acting out situations: a subject acts out how he would say something under specified conditions.

Briere (1971, p. 387) notes that "there is a growing agreement among psycholinguists and sociolinguists that traditional linguistic definitions of the notion of competence in a language are too narrow and are inadequate in identifying all of the skills involved when two people communicate." He concludes that discrete item language tests based on this "narrow definition of linguistic competence" are "inadequate."

Integrative Versus Discrete-Point Items

Proponents of the discrete-item approach maintain the one-question-one-variable posture wherein one and no more than one linguistic element is tested in one item (Lado, 1961; Valette, 1967; Rivers, 1968). A discrete item consists of a short stimulus followed either by an answer to be chosen (a recognition item) or a space for an answer to be given (recall item). Valette (1967) emphasizes the testing of "one thing at a time"; each skill should be tested separately from any other skills, and within each of the language skill tests, only one

point should be tested at a time. Supporters of this type of test item argue that the integrative test item tests several variables simultaneously and depend on judgment on the part of the evaluator (Rivers, 1968; Lado, 1961).

On the other hand, advocates of the integrative-item approach suggest that the discrete-item approach is invalid because of the time allowed for reflection of specific variables that would occur simultaneously in normal conversation (Chastain, 1976). Carroll (in Allen, 1965, p. 369) argues that language testing is not complete without "requiring an integrated facile performance on the part of the examinee. . . . It is possible that knowledge could exist without facility." Herbert (1975, p. 3) agrees and claims that "discrete-item testing can only be an extremely limited sample of the totality of language."

Discrete item loyalists dispute the criticism by pointing out that the rapidity and number of discrete items presented in a short time span compensate; however, Carroll (in Allen, 1965, pp. 369-370) argues that "if we limit ourselves to testing only one point at a time, more time is ordinarily allowed for reflection than would occur in a normal communication situation, no matter how rapidly the discrete items are presented." Carroll recommends tests in which there is less attention paid to specific-structure points or lexicon than the total communicative effect of an utterance.

Test constructors of the integrative test item tradition emphasize the significance of the testing of the creative aspects of language usage. Carroll (in Allen, 1965, p. 370) summarizes the advantages the integrative approach has over the discrete structure-point approach.

> It (the integrative approach) entails a broader and more diffuse sampling over the total field of linguistic items. Furthermore, the difficulty of a task is subjectively more obvious than in the case of a "discrete structure-point item." Thus, when the tasks of an integrative approach test are arranged in the order of their difficulty, the interpretation of performance relative to a subjective standard may be easier.

Robert L. Cooper (in Lugton, 1970, pp. 85-88) adds two dimensions to the original Carroll (in Allen, 1965)-Lado (1961)

testing framework: the parameter of "total linguistic knowledge and the parameter of social aspect--language variety," as illustrated in Figure 4.

Cooper (in Lugton, 1970, p. 86) suggests that the measurement of an integrated skill is better accomplished through the use of direct, integrative items, "involving the simultaneous operation of phonological, syntactic, and semantic knowledge than through the combination of results obtained from subtests or items testing each of these components separately."

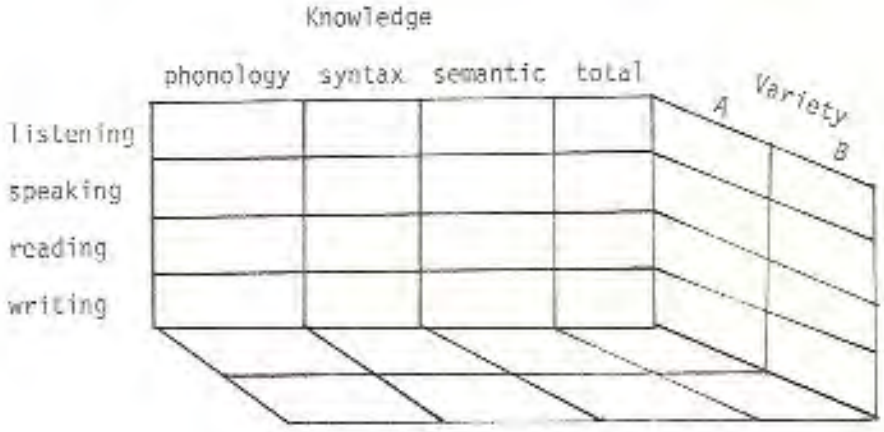

Figure 4. Framework for Specifying Language Test Content. Source: Cooper (in Lugton, 1970, p. 85).

Functional and Criterion-Referenced Tests

Closely aligned with the concept of communicative competence is the concern for "functionalism" in language testing and in criterion-referenced tests. Spolsky et al. (1972, p. 222) suggest that "proficiency tests must be based on a functional definition of levels: tests are needed not of how many items a subject may know (although this is a reasonable question in an achievement or diagnostic test) but of his ability to operate in a specified sociolinguistic situation with ease or effect." Jakobovits (1970) discusses the principles involved in a functional approach to the assessment of

language skills in Foreign Language Learning. In essence, Jakobovits posits that levels of knowing a language cannot be characterized in specific linguistic terms, that is, as mastery of a criterion percentage of items in a grammar and lexicon; the emphasis must be on determining a person's ability to use English as a functional tool of communication. Spolsky et al. (1972) blame the discrete-item approach in proficiency testing for the non-functionalism of most proficiency tests, as explained in a 1968 discussion concerning validity: "The key requirement for discrete-point testing is that we could quantify 'He knows the words on this list'" (Spolsky, 1968, p. 91).

In 1972, Spolsky et al. (1972, p. 220) repeat: "The key assumption of discrete point testing is that it is possible to translate subjective evaluations like 'He doesn't know enough English to understand a lecture' into a series of precise statements like 'he is unable to distinguish between phoneme /i/ and /iy/.'"

Criterion-referenced testing, based on the assumption that it is possible to determine the actual behaviors necessary for adequate performance, was a first attempt to make oral proficiency testing more "functional." Cartier (1968, p. 32) originally suggested the use of criterion-referenced testing in the area of language skills as substitution for the norm-referenced tests which assess "a certain arbitrarily determined percentage of the language." Cartier (1968, p. 29) admits the following problems inherent in such testing:

(1) Criterion testing assumes that a complete and unambiguous inventory can be made of all the behaviors necessary for adequate performance. Linguistic science is not yet sufficiently advanced to provide us with such an unambiguous inventory.

(2) There are no empirically-determined standards of intelligibility, of syntactic accuracy, or of many other aspects of the language which can be applied dogmatically to assessment of a student's capability of performing.

Based on the problems Cartier names, Spolsky (1968, p. 93) posits the need for functional testing of proficiency rather than criterion-referenced testing: "The preparation of proficiency tests . . . would not start from a list of language items, but from a statement of language function." He

goes on to say that the proficiency test should not say "he knows sixty per cent of English," but "he knows enough English to shop in a supermarket." Briere (1971, p. 388), however, criticizes the notion of criterion-referenced testing when he asks, "How does one go about deciding when someone knows enough language to carry out a specific function?" Briere suggests characterizing the linguistic knowledge which correlates with the functional ability and then testing that language.

Overall Proficiency

Harry L. Gradman and Bernard Spolsky (in Jones and Spolsky, 1975, p. 67) contend that there exists something that they have labeled "overall proficiency." They base this contention on a "theoretical notion of knowledge of a language and the assumption that while this knowledge at a certain level can be divided up into various kinds of skills, there is something underlying the various skills which is obviously not the same as competence" (Jones and Spolsky, 1975, p. 67). Briere (1971, p. 387) describes "overall proficiency" as analogous to the unseen iceberg underneath the surface, "The part of the iceberg seen floating on top of the water is but a small fraction of what lies underneath the water." Spolsky (in Jones and Spolsky, 1975) maintains that we must find ways of getting beyond the limitations of testing a sample of surface features--to tap the underlying linguistic features.

Current Experimental or Working Tests

Oral Interview

The interest in the use of the oral interview as the most appropriate test for the assessment of oral proficiency reflects the influence of the aforementioned factors of communicative competence, integrative test items, and functionality.

Long considered to be too unreliable to be valid, too cumbersome, too time-consuming, and too difficult to score because of the integration of the skills (Lado, 1961; Valette, 1967; Rivers, 1968), the oral interview because of its "real-life" quality (Clark, in Jones and Spolsky, 1975), is now considered to be the most direct kind of measure of communicative competence thus far developed. According to John L. D. Clark

(in Jones and Spolsky, 1975, p. 11), "a direct test of oral proficiency, in the face-to-face communication sense, would involve a test setting in which the examinee and one or more human interlocutors do, in fact, engage in communicative dialogue." Lado, on the other hand, disagrees with the merit in the face validity of the oral interview, and criticizes the oral interview for these reasons:

(1) The interviewee cannot ask questions.

(2) The interviewee is under a tremendous tension during the interview situation.

(3) The examiner may allow other than related variables to influence his scoring, such as weariness, etc. (in Discussion in Jones and Spolsky, 1975, p. 25).

His criticisms, however, are countered by Clark (Discussion in Jones and Spolsky, 1975, p. 27) by reminding Lado that the purpose of the interview is to create a "natural" setting.

An example of an uncontrolled oral interview is the one used by the Foreign Service Institute of the Department of State as described by Claudia P. Wilds (in Jones and Spolsky, 1975, pp. 29-38). Used for the purpose of rating the language proficiency of government employees, each interview is structured to accommodate the examinee's experience and capabilities in the language.

The testing team at the Foreign Service Institute consists of a native speaker of the language being tested and a language examiner who may be an experienced native-speaking language instructor or a linguist thoroughly familiar with the language. At the Central Intelligence Agency, two native speaking language instructors conduct the test, whereas at the Foreign Language Institute, although two examiners are present, only one conducts the test while the other observes and records.

The test begins with customary social exchanges including introductions, comments about the weather, and questions like "Have you just come back from overseas?" The examinee who has difficulty answering the preliminaries even with repetition and rephrasing will be asked minimum-difficult questions about himself, his family, or his work. An examinee who can perform at this level is usually rated between S-1 and S-2 (see Rating Scale, Figure 5).

S-1	Elementary Proficiency. The examinee is able to satisfy routine travel needs and minimum courtesy requirements.
S-2	Limited Working Proficiency. The examinee is able to satisfy routine social demands and limited work requirements.
S-3	Minimum Professional Proficiency. Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics.
S-4	Full Professional Proficiency. Able to use the language fluently and accurately on all levels pertinent to professional needs.
S-5	Native or Bilingual Proficiency. Speaking proficiency equivalent to that of an educated native speaker.

Figure 5. Proficiency Categories

On the other hand, the examinee who successfully answers the preliminary questions is led into natural conversation on autobiographical and professional topics. The experienced interviewer will simultaneously try to elicit the grammatical features that need to be checked. As the questions increase in complexity and detail, the examinee's limitations in structure, vocabulary, and comprehension become apparent. If the examinee avoids the complex constructions, the raters will use an informal interpreting situation to further test the examinee, which comprises the second part.

The third portion of the speaking test involves instructions or messages written in English and given to the examinee to be communicated to the native speaker, such as "Tell your landlord that the ceiling in the living room is cracked and leaking and the sofa and rug are ruined" (Wilds, in Jones and Spolsky, 1975, p. 31).

The examinees are rated according to the performance factors in Figure 5 and then placed into proficiency categories illustrated in Figure 6 (Wilds, in Jones and Spolsky, 1975, pp. 29-38).

1. Accent	foreign	___	___	to	___	___ native
2. Grammar	inaccurate	___	___	to	___	___ accurate
3. Vocabulary	inadequate	___	___	to	___	___ adequate
4. Fluency	uneven	___	___	to	___	___ even
5. Comprehension	incomplete	___	___	to	___	___ complete

Figure 6. Checklist of Performance Factors

What Clark (in Jones and Spolsky, 1975, p. 11) classifies as indirect oral proficiency tests are oral interviews which "involve certain quasi-realistic activities on the students' part." He includes in this category oral interviews based on pictures, such as the Ilyin Oral Interview. Ilyin (1975, p. 9) refers to an uncontrolled interview as "guesswork that poses as an oral interview." She considers the Ilyin Oral Interview with its consistent procedural and scoring pattern, as a useful tool in objectively assessing oral levels of proficiency.

The interview requires only one examiner per examinee. Ilyin (1975, p. 9) states that "any friendly, relaxed person who understands the test and can read in a conversational way can give the test; and any educated objective person can score it. Many paraprofessionals and aides have been highly successful in giving the interview." Based on a set of pictures as stimuli which are divided into past, present, and future categories, the interview consists of the elicitation of questions by both examiner and examinee, who is asked to answer in complete sentences. The responses of the examinee are manually recorded and scored at a later time, according to the following criteria: information, word order, verb structure, and other. Examiners score the response only when appropriate information has been communicated. The examiner does not penalize for accent if the information is understandable in context, and he accepts any variations in structure and/or

grammar that educated native speakers would use in the context. Figure 7 is a sample set of four items and how they were scored, according to a rating scale of 0-1 (Ilyin, 1975, p. 15).

E = Examiner
S = Student

Items	Experimental Edition (Long System)			
		Accuracy		
	Info	Word Order	Verb	Other
8) E: How does Bill go to school?				
S: He go walking to school.	1	1	0	1
9) E: Where does he eat his lunch on weekdays?				
S: They eat after school	0	0	0	0
10) E: When does he eat lunch on weekends?				
S: Is at 11:50 minutes to 12 and a half	1	1	1	0
11) E: Is he going to be eating lunch tomorrow at 12:15?				
S: Oh, no.				
E. Good. Tell me in a complete sentence.				
S: No, he isn't	1	1	1	1

Figure 7. Sample from Ilyin Oral Interview -- Source: Ilyin (1975, p. 15).

Indirectly criticizing Ilyin's Oral Interview, Herbert (1976, p. 4) states that an examinee should not be placed in the examiner role because the role would tend to threaten the examinee thereby limiting or distorting his natural responses. Herbert (1975) is currently experimenting with an oral test based on no elicitation of language by the examiner; instead the examinees (children, in this case) are shown various pictures and objects; student responses are mechanically recorded. The sample is then graded according to criteria involving sentence complexity.

Functional Tests

Spolsky et al. (1972) devised three functional tests of oral proficiency. A description of two of the tests follows:

1. Spanish-English Language Dominance Assessment. The assessment was developed for use with first and second grade pupils. The test was based on the following assumptions:

   a. While a school system needs a single decision, bilingual dominance varies from domain to domain. Subscores were therefore given for the domains of home, neighborhood, and school.
   b. A child's report of his own language use is likely to be quite accurate.
   c. Vocabulary fluency (word naming) is a good measure of knowledge of the language and it is a good method of comparing knowledge of two languages.
   d. The natural bias of the schools in Albuquerque as a testing situation favors the use of English; this needs to be counteracted by using a Spanish-speaking interviewer.

   The interview falls into three main sections, each with a Spanish and English component: (a) a language background questionnaire, (b) word-naming tasks, and (c) description of pictures. The following are the categories which provide some guidelines for assessment:

   S-S: Spanish dominant. The child's ability in English varies, but he uses Spanish more easily.
   S-E: The child seems to handle each language equally easily.

E: English dominant. There is variation in the child's ability to understand and speak Spanish but he seems most at home with English.
E: English monolingual.
N: The child did not respond sufficiently for any judgment to be made.

2. "The Oral Placement Test for Adults" is intended to decide whether an adult who lacks literacy in either English or Spanish but speaks Spanish, needs instruction in English before going on to regular adult basic education. The interview consists of a number of blocks. The first block consists of seven questions about the examinee. If the student cannot reply, he is placed at level one. The second block consists of 14 simple sentences to be repeated with an understandable pronunciation; the "assumption is that the person who can repeat these items with such underlying structural differences is better able to function in English than one who cannot repeat them" (Spolsky et al., 1972, p. 230). In a summary statement, Spolsky et al. (1972, p. 231) conclude that "these three tests display certain common principles in developing useable functional tests. Each is closely tied to the practical situation for which it was prepared, and is intended to be used by relatively untrained testers with the simplest possible materials."

The blocks 3-14 increase in structural complexity and in the complexity of the manipulation called for: i.e., negative and question transformation and control of tense structures.

Madsen's (1974) test, developed for the Migrant Education Program of California, is a functional test of performance. His test consists of a number of commands to which the examinee must respond to in action and speech. For example, the examiner tells the examinee "to pick up the phone and call" a number and give a particular message to the person who answers. The test is functionally oriented in that the degree of English proficiency the examinee needs for a work situation is the underlying critical question. The greatest problem with this kind of test is the cultural variable that it might account for rather than the communicative one. For the population for which

the test was intended, it is doubtful that it is either reliable or valid, but for another test population it might prove to be very appropriate.

Tests of Overall Proficiency

A very recently developed indirect test of overall proficiency is the reduced redundancy procedure developed by Spolsky et al. (1968) and further refined and developed by Gradman and Spolsky (in Jones and Spolsky, 1975). The procedure involves dictating to the examinee a number of sentences in the target language which have been distorted by the introduction of "white noise,"* at various signal/noise levels, whereupon the student attempts to write out each sentence as it is heard. Gradman (in Jones and Spolsky, 1975, pp. 67-68) defines language redundancy as the "variety of structural and lexical clues in a sentence which may be obscured by adding noise to the background." The message, however, may still be understood by a native speaker or by near-native speaker. According to Spolsky (in Jones and Spolsky, 1975), the test measures overall proficiency. The assumption on which the test is based is that students who have a high degree of overall language proficiency can continue to understand the recorded sentences even when many linguistic cues have been destroyed. The sentences are based on five sentence types: simple negative questions, simple questions, simple passives; a category called embedded, embedded negatives, embedded questions, embedded questions signaled by intonation only, embedded negative questions, and a category called miscellaneous. Spolsky (in Jones and Spolsky, 1975, pp. 62-63) reports his test to have a .60 correlation with the Ilyin Oral Interview and the TOEFL total score. However, the general criticism is that the test might have the ability to distinguish native speakers from non-native speakers, but that it has trouble establishing gradations of proficiency (Spolsky, in Jones and Spolsky, 1975, p. 70).

---

* White noise is comprised of the random frequencies at random amplitudes, the basic kind of noise that is heard in the background of radio broadcasts. It is called white because it has the same characteristics as white light, that is, all frequencies are represented at random (Jones and Spolsky, 1975, p. 62).

Another method of testing overall language proficiency currently in use is the "cloze" test. The technique, originated by Taylor (1953) in the context of native-language testing, consists of the systematic deletion of letters or words from a continuous printed text. The student is asked to supply the missing item on the basis of contextual clues available in the remaining portion of the text. Numerous experimental studies have been conducted (Carroll, Carton, and Wilds, 1959; Oller and Conrad, 1971), including the investigation of deleting only certain categories of words, such as prepositions (Oller and Inal, 1971); computer-based scoring using a "clozentropy" formula based on information theory (Darnell, 1970); and the acceptance of contextual variants, not necessarily the originally deleted word (Oller, 1972). Darnell (1970, p. 36) calls his cloze test "an ideal test of proficiency which will measure the adequacy of an individual's linguistic performance rather than his awareness of linguistic rules or his ability to answer questions about the language."

Darnell (1970, p. 44) suggests the following advantages of the clozentropy test:

1) It is based on a functional, rather than a formal definition of proficiency.

2) It is an objective procedure which may be scored by computer.

3) It is adaptable to the needs of specialized groups and specialized uses of language. It is logically as appropriate to measure one's command of a jargon or a particular local dialect as it is to measure control of general American.

Both reduced redundancy tests and cloze procedures offer considerable advantages; however, the ultimate usefulness of these and other indirect techniques of measuring oral proficiency depend on the correlations that can be developed between them and the more direct measures of oral proficiency such as the oral interview.

Conclusion

In his article entitled "Theoretical and Technical Considerations in Oral Proficiency Testing," John D. Clark (in Jones and Spolsky, 1975, p. 23) concludes the following:

> If oral proficiency is defined as the student's ability to communicate accurately and effectively in real-life language-use contexts, especially in the face-to-face conversations typical of the great majority of real-world speech activities, considerations of face validity appear to require human administration of a conversation-based test, which must also be evaluated by human raters.

He both commends and recommends the further testing and development of oral interview techniques on the basis of their high face validity. He encourages the investigation of techniques which will assure acceptable reliability levels within short testing times and which will improve scoring procedures.

REFERENCES

Allen, Harold B. Teaching English as a Second Language. New York: McGraw-Hill Book Company, 1965.

Briere, E. L. "Are We Really Measuring Proficiency with Our Foreign Language Tests?" Foreign Language Annals, Vol. 4, 1971, pp. 385-391.

Buiten, Robert, and Harlane Lane. "A Self-Instructional Device for Conditioning Accurate Prosody," IRAL, Vol. 3, 1965, pp. 205-219.

Cartier, Francis. "Criterion-Referenced Testing of Language Skills," TESOL Quarterly, Vol. 2, March 1968, pp. 27-32.

Chastain, Kenneth. Developing Second Language Skills: Theory to Practice. 2nd ed. Chicago: Rand McNally College Publishing Co., 1976.

Darnell, D. K. "Clozentropy: A Procedure for Testing English Language Proficiency of Foreign Students." Speech, Vol. 37, March 1970, pp. 36-46.

Fries, Charles. Teaching and Learning English as a Foreign Language. Ann Arbor: University of Michigan Press, 1953.

Gumperz, John, and Dell Hymes. Directions in Sociolinguistics. New York: Holt, Rinehart, and Winston, 1970.

Harris, David P. Testing English as a Second Language. New York: McGraw-Hill, 1969.

Herbert, Charles H. "Language Diagnosis and Assessment and a Program for Natural Language Development." Paper presented at 1975 TESOL Convention, Los Angeles, 1975.

Ilyin, Donna. "Assessing Oral Communication in Adult Program English Second Language Classes." Paper presented at the Annual Convention of the Teachers of English to Speakers of Other Languages, 1975.

Jakobovits, Leon A. Foreign Language Learning. Newbury House, 1970.

Jones, Randall L., and Bernard Spolsky. Testing Language Proficiency. Arlington, Virginia: Center for Applied Linguistics, 1975.

Lado, Robert. Language Testing. New York: McGraw-Hill Book Company, 1961.

Lugton, Robert C. Preparing the EFL Teacher: A Projection for the 70's. Philadelphia: The Center for Curriculum Development, 1970.

Madsen, Harold S. "An ESL Performance Examination." Paper delivered at the Eighth Annual TESOL Convention, Denver, Colorado, March 1974.

Oller, John W., Jr. "Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language." Modern Language Journal, Vol. 56(3), 1972, pp. 151-158.

Paulston, Christina Bratt. "Linguistic and Communicative Competence." TESOL Quarterly, Vol. 8, December 1974, pp. 347-362.

Pulliam, Robert. Application of the SCOPE Speech Interpreter in Experimental Educational Programs. Fairfax, Virginia: Pulliam and Associates, 1969.

Rand, Earl. "A Short Test of Oral English Proficiency." Language Learning, Vol. 13, 1963, pp. 203-210.

Rivers, Wilga. Teaching Foreign Language Skills. Chicago: University of Chicago Press, 1968.

Rivers, Wilga. "Talking Off the Tops of Their Heads." TESOL Quarterly, Vol. 6, March 1972, pp. 71-81.

Spolsky, Bernard. "Language Testing--The Problem of Validation." TESOL Quarterly, Vol. 2, June 1968, pp. 88-94.

Spolsky, Bernard, Penny Murphy, Wayne Holm, and Allen Ferrell. "Three Functional Tests of Oral Proficiency." TESOL Quarterly, Vol. 6, September 1972, pp. 221-236.

Taylor, Wilson L. "Cloze Procedure: A New Tool for Measuring Readability." Journalism Quarterly, Vol. 30(4), 1953, pp. 414-438.

Valette, R. M. Modern Language Testing: A Handbook. New York: Harcourt Brace and World, 1967.

White, Ronald V. "Communicative Competence, Registers, and Second Language Teaching." IRAL, Vol. 7, May 1974a, pp. 127-141.

The Mextesol Journal wishes to thank the English Language Teaching Journal for permission to print the article, "The Overhead Projector in the Advanced Reading Class," by Miss Janet McAlpin. This article will appear in the English Language Teaching Journal, Vol. XXXIII, No. 3, April 1979.