DISCRETE STRUCTURE-POINT TESTING: PROBLEMS AND ALTERNATIVES

by Kenneth G. Aitken

This paper is intended to present some reasons for reconsidering the use of discrete structure point tests of language proficiency. It is also intended to suggest an alternative basis for designing proficiency tests.

Perhaps to begin it would be useful to define some terms used in language testing. Language tests have been divided into four categories based on their use. Language aptitude tests are used to predict probable success or failure in certain kinds of language study. Language achievement tests are used to determine how effective teaching has been, or how much of what has been taught has been learned. Diagnostic tests point out areas in which a student requires additional concentrated teaching and study. Language proficiency tests indicate whether or not an individual is proficient enough in a language to perform certain tasks or undertake certain training programs in the target language. As we previously stated, the focus of this paper is on proficiency testing.

John B. Carroll (1961) has divided proficiency tests into two types: "discrete structure point" (or "discrete point") and "integrative" tests. "Discrete point" tests are based on the assumption that there are a given number of specific structure points, the mastery of which constitutes "knowing" a language. On the other hand an "integrative" test is one based on the premise that "knowing" a language must be expressed in some type of functional statement such as, 'He knows enough English to read the paper, but not enough to follow the news on television!' (Spolsky 1968).

Discrete point tests tap those areas that linguists include in linguistic competence: vocabulary knowledge, recognition of correct grammatical structures, sound discrimination, etc.—in other words, the mechanics of the language. Integrative tests, Jakobovits (1970) contends, tap communicative competence factors—i.e. those factors crucial to how a language is used for communicative purposes.

There is reason to believe that the discrete point approach to language proficiency testing, one of the primary tools of disciples of the audio-lingual method of teaching a foreign language, is based on erroneous assumptions. The discrete point approach is fundamental to the approach taken by Fries (1952) and Lado (1957) who pioneered the audio-lingual method.

The discrete point approach to language teaching and testing rests on both traditional verbal learning theory (Dulay and Burt, 1972) and American structuralist linguistic theory (Oller 1973 a). The basic assumptions are:

(1) The surface structure of a language can be systematically described, and its elements listed and compared with any other language, similarly described.
(2) The mastery of a language may be divided into the mastery of a number of separate skills; listening, speaking, reading, and writing. These skills in turn may be divided into a number of distinct items. It assumes that to have developed a criterion level of mastery of skills and items listed for that language (e.g. 50,000 discrete items) is mastery of the language.
(3) The contrastive analysis hypothesis, (which states that if language learning is habit formation, where L2 is similar to L1 there will be a positive transfer of learning, and where L2 and L1 differ there will be a negative transfer: i.e. mother tongue interference) is true.

It has been convincingly argued by others that on theoretical and empirical grounds the contrastive analysis hypothesis is unsound (Dulay and Burt 1972, Upshur 1962, and Wardhaugh 1970). In this paper we will focus our attention on assumptions (1) and (2) as these ideas continue to prevail while the number of adherents to the CA hypothesis dwindles.

There are serious theoretical objections to assumption (1) because it assumes that knowledge of a language is finite in the sense that it is possible to make a list of all its items. However, if such an impossible task were completed, and a discrete point test constructed based on a representative sample of the language the test would be too long to be practical. On the other hand anything less than a representative sample of the total language would significantly affect the validity of the tests as a determiner of language proficiency. The criterion for composing the representative sample would have to be on the basis of functional necessity, which involves defining the functional load of ability, for example, to recognize the appropriateness of a given verb form. Spolsky (1968 b) argues that to do this we would have to collect a list of minimal pair utterances in which the distinction is vital: that is, where a single linguistic difference in a given situation will lead to complete misunderstanding--an extremely rare situation. The rarity of such a situation is the result (and theoretical cause) of the redundancy of natural languages.

The second assumption is, according to Spolsky (1968 b), crucial to this approach to testing:

> The key assumption of the discrete-point approach is that it is possible to translate sentences. . . .(like) "He doesn't know enough English to write an essay, but he seems to be able to follow lectures and read his textbooks without much trouble.". . . .into a list of sentences. . . . .(like) "He is unable to distinguish between the phonemes /i/ and /iy/.",. . . .The key requirement of discrete point testing is that we could quantify"He knows the words on this list." (1968 b:92)

Research by Spolsky et al. (1968) and other psycholinguists has isolated the ability to utilize the redundancy of natural language as a factor in "knowing" a language even in situations of reduced redundancy. Redundancy is the property of natural languages that allows one to predict missing elements from the context (Carroll 1964:44-65). Highly redundant messages tend to be repetitive and to contain relatively little information per element. By "information" is meant the informativeness of the elements (Shannon and Weaver 1949). Since the redundancy factor in natural language allows us to predict missing elements in a message or utterance, then the greater the redundancy, the greater the chance the message will be comprehended. A model of understanding speech must then include the ability to make valid guesses about a certain percentage of unknown items in the language.

Kenneth G. Aitken, a former BATESL major at Church College of Hawaii (now Brigham Young University-Hawaii Campus) is an ESL instructor at Vancouver Community College in Vancouver, B.C., Canada, where he is working on the testing and placement of ESL students.

Spolsky (1968 a) contends that this raises some serious theoretical questions about the value of deciding a person knows a language because he knows certain items in the language, i.e. assumption (2). The principle of redundancy suggests that it will not be possible to demonstrate that any given language item is essential to successful communication, nor to establish the functional load to any given item in communication as discussed previously.

Looking at it from another angle, we note that assumption (2) has roots in the hypothesis that the meaning of a linguistic form is defined as the situation in which the speaker utters that form and the response it calls forth in the hearer (Bloomfield 1933:139). Oller, Sales and Harrington (1969:318) argue that this definition is both misleading and inadequate. The meaning of a linguistic unit in any communicative situation, they state, cannot be adequately construed as a given speaker's situation and hearer's response; rather it is dependent on (1) at least one individual's history of experience with that unit in other situations, and (2) the inductive-deductive

processes which enable that individual to relate the information from his history of experience to the situation in question. This definition places a premium upon the role of one's personal experience bank in language use. Assumption (2) does not take into account the role of previous personal experience in the assignment of meaning in language use. Counter evidences to this assumption are utterances like "Why don't you wipe that silly grin off your face," which rely heavily upon the listener's previous personal experience to provide him with the criteria for an appropriate response.

As Belasco (1969) has observed, many points of grammar all well learned, but learned separately and unrelatedly do not constitute proficiency in a language. In fact, they cannot in principle. Language is essentially different in type from any list of discrete items. The essence of discrete point fallacy, argues Oller (1973), is the incorrect assumption that a test of many isolated and separate points of grammar or lexicon is a test of language in any realistic sense.

It seems that the discrete point approach to objectively test language proficiency is not the pedagogical panacea many language teachers had assumed. It is unfortunate that so many have made the much stronger claim that the discrete point approach is the best basis for language test design.

Having examined some of the theoretical problems relating to the discrete structure point approach to testing, we shall turn our attention to the integrative test approach. Our object, we must remember, is to develop an approach to testing that can be used in making proficiency examinations. Proficiency tests, we defined as being used to tell us whether or not an individual is proficient enough in a language to perform certain tasks, or to undertake a certain training program in that language. Our previous discussion suggests the impossibility of determining language proficiency in linguistic terms. A more promising approach might be to work for a functional definition of levels of proficiency. In other words, we do not aim at how much of a language the student knows, but rather, we test his ability to operate in a specified sociolinguistic situation with specified ease or effect.

Spolsky (1968 b) suggests that starting with functional statements of this sort, the language tester's task is to find a reliable, valid, and economical method of rating a student's proficiency in these terms. He proposes a two stage research strategy. First, ignoring costs in time and money, the tester should have the subject evaluated by trained panels of judges in situations of the sort described functionally in the rating scale. Then these evaluations become the yardstick to which more economical measures are correlated. The degree of correlation will reflect the value of the ad hoc tests and make clear the degree of doubt that must be kept in their interpretation.

Oller (1973 b), realizing that language teachers cannot wait for linguists to solve all the problems and theoretical issues involved, recommends several tests of integrative skills, especially cloze tests and dictation, both types of tests requiring control of the natural redundancy of language. Another integrative test involving error recognition was recommended by Sibayan (1971).

Cloze tests (Taylor 1973), originally conceived as reading comprehension and/or readability measure, are constructed by systemetically deleting every n-th word from a prose passage. In foreign languages, every seventh word is usually deleted over a passage of about 375 words, leaving the first and last sentences intact. The subject must read the mutilated passage and, based on his comprehension of it, and his language experience, guess the deleted words. The score is based on the percentage of correct guesses. Another scoring method involves accepting any contexually acceptable answer. Answers in both scoring methods must not violate any grammatical restrictions. Oller (1973 c) and his reference provide an excellent background and review of the use of cloze in ESL, including several validation studies.

Cloze tests can be constructed that are criterion based in relation to the situation in which the students are being trained to function (Aitken 1975). For example, if we were preparing a group of foreign engineering students to enter a post-graduate program in civil engineering, the prose passages we used to create the cloze test might be chosen from

(continued on page 18)

Discrete Structure Point Testing

(continued from page 9)

current journal literature in the field of civil engineering. On the other hand, if the group is comprised of new immigrants who are studying English and plan to become citizens, perhaps, having discussed national geography, government, culture and history in their lessons, we might select a passage from some of the appropriate literature on these topics for the cloze test.

Dictation as a test has had its ups and downs in popularity among professional language test writers. After many years of relative unpopularity, dictation is now being recognized as an effective diagnostic device for the classroom teacher, and an excellent integrative test to include in a language proficiency test battery (Oller 1973 b). Of course, many of us knew it all the time.

Dictation passages may be presented in a number of ways. Valette (1967) recommends one effective technique of administering a test dictation.

> "First, the whole passage is read at normal speed. The students are told not to write, just to listen carefully. Then the passage is read a phrase at a time, with pauses during which the students write down what they have heard. At this time the teacher may read each phrase either once of twice, as long as he is consistent. (At the teacher's discretion, punctuation marks may be given in the target language.) Finally, the entire passage is read again at normal speed, and the students are given a few minutes for final revision. It is imperative that the teacher never repeat a particular phrase at a student's request." (p. 140)

Scoring the dictation may be done in various ways. One method is to start with a total score of 20 and deduct half a point for each spelling error and one point for each other error. Another method would be to give one point for each word correct in the passage transcribed. A third method would be to give a certain number of points per phrase transcribed correctly.

Typical student errors in taking dictation include omitting words, using the wrong word, putting the words in the wrong order, and using ungrammatical phrases and sentences. For further discussion the reader is referred to Oller (1973 b) and his references.

Like the cloze passages, dictation passages can be criterion related; that is chosen from any appropriate context in which the students are preparing to function. Dictations of lists of items, or short chunks of data, as suggested by some teachers, may be appropriate for testing comprehension in some areas, but overall proficiency is probably best tested with a passage dictation.

Error recognition tests of the type recommended by Sibayan (1971) are somewhat more difficult to construct than the other two integrative tests, yet they are still simple enough that valid and reliable items can be constructed by an ESL teacher with little experience in test construction. Each item consists of four choices, the first three being statements in which grammatical elements, capitalization and punctuation are critical. The fourth choice is (D) No mistakes. The directions to the student are as follows:

> Each item below contains three sentences. These sentences may be correct, or one of them may have a mistake in grammar, punctuation, or capitalization. If all the sentences are correct, mark (D) on your answer sheet. If one of the sentences contains an error, mark the letter of that sentence on your answer sheet.

There are two ways to construct the choices in the item. One could choose any three potentially difficult sentences for the first three choices. However, if one were to make the three sentences into a continuous thought progression, the degree of integration of language skills would be extended. The following test item is an example of this type of integration.

A. Since Jack was five years old he has played the piano.
B. Jack was hoped to become a famous musician.
C. Unfortunately, Jack never learned how to read music.
D. No mistakes.

Harris (1969:73) recommends error recognition tasks, combined with sentence completion tasks, as valid objective measures of writing ability. To the extent that cloze tests are sentence completion tasks, the combination of cloze and error recognition tasks is potentially a powerful proficiency test. However, a search of the literature on second and foreign languge testing reveals no published research on error recognition tasks of the above mentioned type. At this point there appears to be a need for extensive research in various types of error recognition tasks in ESL testing.

> "In production of language, expectations are formed as to what the output should look like, and the output is modified until it conforms to the expectations. In reception of language, expectancies are constantly generated as to what is likely to follow in a given sequence, and are modified to match received input. The first process is actually one of synthesis-by-analysis and the second is one of analysis-by-synthesis. In both cases there is a matching of physically realized signals to expectancies generated on the basis of an underlying grammar." (p. 8)

We can contruct a test that does not separate language from its use, allowing us to test communication in contexts similar to, if not the same as contexts the students would normally encounter in the real world. The real world outside our language classrooms is not concerned particularly whether the language skills are refined to the n-th degree, as much as it is concerned about how effectively one employs the vast number of linguistic, paralinguistic and non-linguistic strategies available to the individual in the context of the communication act.

Oller (1973 a) proposes a theoretical base for integrative tests. In any normal use of language, he explains that the sequence of linguistic elements that occur is restricted by various kinds of context.

To sum up, we have tried to suggest some of the weaknesses of the current trend of using discrete point tests to determine language proficiency in second languages. The discrete point basis has been falsified by revealing the theoretical flaws in its foundation. A functional approach based on effective language use in appropriate situations has been proposed. Finally Oller's theory of a grammar of expectancy, which incorporates the redundancy utilization factor, has been introduced to provide a theoretical base for an integrative approach to testing.

It remains to be seen whether any institute or body of applied linguists will develop a scale of functional language use as proposed by Spolsky. It also remains to be seen if Oller's pragmatic approach—grammars of expectancy—has been defined in theoretical terms that can be falsified, and if they are so defined, whether they can be falsified. These problems are beyond the scope of this present paper.

REFERENCES

Aitken, Kenneth George. (1975)
"Problems in Cloze Testing Re-examined," *TESL Reporter*, vol. 8, no. 2.

Bloomfield, Leonard. (1933)
Language, Toronto: Holt Rinehart and Winston, 1933.

Briere, Eugene J. (1971)
"Are We Really Measuring Proficiency with Our Foreign Language Tests?" *Foreign Language Annals*, vol. 4, no. 4 (May 1971), pp. 385-391.

Carroll, John B. (1964)
Language and Thought. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Carroll, John B. (1961)
"Fundamental Considerations in Testing English Proficiency of Foreign Students," *Testing*, Washington, D.C.: Center for Applied Linguistics, 1961, pp. 31-40.

Fries, Charles C. (1945)
Teaching and Learning English as a Foreign Language, Ann Arbor: University of Michigan Press.

Harris, David P. (1969)
Testing English as a Second Language. Toronto: McGraw-Hill.

Holtzman, Paul D.
"English Language Proficiency Testing and the Individual" in *Selected Conference Papers of the Association of Teachers of English as a Second Language.* (Washington): NAFSA, 1967.

Jackobovits, Leon A. (1970)
Foreign Language Learning: A Psycholinguistic Analysis of the Issues. Rowley, Mass.: Newbury House, 1970.

Lado, R. (1957)
Linguistics Across Cultures. Ann Arbor: University of Michigan Press, 1957.

Oller, John W. Jr. (1973 c)
"Cloze Tests of Second Language Proficiency And What They Measure," *Language Learning,* vol. 23, no. 1 (June 1973), pp 105-118.

Oller, John W. Jr. (1973 b)
"Discrete-point Tests Versus Tests of Integrative Skills," in Oller and Richards (eds.), *Focus on the Learner,* Rowley, Mass.: Newbury House, 1973.

Oller, John W. Jr. (1974)
Expectancy for successive elements: key ingredient to language use, *Foreign Language Annals,* vol. 7 (March 1974), pp. 443-452.

Oller, John W. Jr. (1973 a)
Pragmatic language testing, *Language Sciences,* no. 28 (December 1973), pp. 7-12.

Oller, John W. Jr.; B. Dennis Sales and Ronald V. Harrington. (1969)
A basic circularity in traditional and current linguistic theory, *Lingua,* vol. 22 (1969), pp. 317-327.

Shannon, C.; and W. Weaver, (1949)
The Mathematical Theory of Communication, Urbana: University of Illinois Press, 1949.

Sibayan, Bonifacio. (1971)
"Elements to Test in English as a Second Language," in R. Cherrier (ed.). English *Language Testing: Report of the RELC Fifth Regional Seminar Bangkok 25-30, May, 1970.* Singapore: Regional English Language Centre.

Spolsky, Bernard. (1968 b)
"Language Testing—the Problem of Validation," *TESOL Quarterly,* vol. 2 (1968), pp. 88-94.

Spolsky, Bernard; Bengt Sigurd; Masahito Sato; Edward Walker and Catherine Aterburn. (1968)
"Preliminary Studies in the Development of Techniques for Testing Overall Second Language Proficiency, *Language Learning,* Special Issue no. 3 (August (1968), pp. 79-101.

Spolsky, Bernard. (1968 a)
"What Does it Mean to Know a Language, or How Do You Get Someone to Perform His Competence?" Paper presented at the Second Conference Problems in Foreign Language Testing. University of Southern California, November 1968.

Taylor, Wilson S. (1953)
"Cloze Procedure: A New Tool for Measuring Readability. *Journalism Quarterly,* vol. 30 (1953), pp. 415-433.

Upshur, John A. (1962)
"Language Proficiency Testing and the Contrastive Analysis Dilemma." *Language Learning,* vol. 12, no. 2, pp. 123-127.

Valette, Rebecca. (1967)
Modern Language Testing: A Handbook, New York: Harcourt, Brace and World, 1967.

Wardhaugh, Ronald. (1970)
The Contrastive Analysis Hypothesis, *TESOL Quarterly,* vol. 4, no. 2 (June), pp. 123-130.

TESL REPORTER
Box 157
Brigham Young University
Laie, Hawaii 96762

Non-Profit Organization
U.S. POSTAGE
PAID
Laie, Oahu, Hawaii
Permit No. 1