

English Language Testing in Korea

- With Special Respect to the College Scholastic Ability Test

Byoung-Man Jeon*

1. Introduction

When we consider the current situation of language testing in such a test-dependent society as Korea, we find that the College Scholastic Ability Test (CSAT), one of the most important and crucial tests, is considered 'backwash', which can be beneficial to language teaching/learning methods. Until 1993, the English tests for College Entrance had not appropriately measured the four skills of English but only one skill, reading comprehension ability. Moreover, most of the test items used in these tests are multiple-choice (MC) ones, which are useful for maintaining the scientific accuracy because of the massive number of the examinees. However, trends and tendency of language testing today should be direct and authentic.

In this paper, I'll briefly review not only the historico-theoretical perspective examining the development of language testing but various test types such as Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS). Also, I'll review the development of the English tests for College Entrance Examination such as the

* Chonbuk Nat'l University

Preliminary Test for the College Entrance (PTCE), the Scholastic Ability Test (SAT), and the College Scholaristic Ability Test (CSAT). Finally, I'll discuss the problems of English tests of these three test types used in Korea and suggest the desirable directions for appropriately using the CSAT in the Korean educational situation.

2. Overview of Language Testing Theories

2.1 Development of Language Testing

In reviewing the development of language testing, Spolsky (1978) suggested three major historical periods in language testing: 'prescientific' period, 'psychometric-structuralist' period, and 'integrative-sociolinguistics' period. The first period was the one roughly prior to the early 1950s. During this period, the main teaching method was grammar-translation, i.e., reading oriented teaching approach, through which the learners of foreign language could improve their reading ability by practicing translation between the target language and the native one. Moreover, this method included the exercises of grammar concerning special usages, exceptions to the general rules, etc. This method was pointed out to have many drawbacks in dealing with functional proficiency of communication because it focused primarily on the written language instead of the spoken one.

Language tests consisted largely of translation or measurement of knowledge about grammatical items. As a result, these kinds

of tests depended entirely on the teacher's subjective judgment and thus rarely had the scientific analysis such as validity, reliability, item analysis, etc. The following quotation from Clark (1983: 431) explains the criticism of language testing of this period:

...the testing approaches utilized during this period would obviously not for the most part have served to measure validly the student's ability to make use of the language as a means of functional communication in 'real-life' situations outside of the classroom setting.

Next, the psychometric-structuralist period extended from the early 1950s to the late 1960s, and it was closely related with structuralism and empiricism. During this period, the most striking of the influential methods was contrastive analysis introduced by Lado (1961), which compared the target language with the native one. Lado extended this assumption to language testing by asserting the necessity of discrete point testing. In order to maintain the scientific accuracy of language testing, discrete point testing was developed as a measurement to test one point at a time. Especially it was attempted that the selection of test items should be based on contrastive analysis which was considered a more reliable scientific approach than the teachers' voluntary intuition.

Language testing during this period tended not only to measure the correlation coefficient between test scores but also to validate the item analysis through the scientific method. However, though these methods also had many

advantages over those of the pre-scientific period, Clark (1983: 432) mentioned the weak points as following:

By its nature, however, the discrete-point, one-element-per-item testing procedure was not capable of measuring the student's ability to comprehend or produce, on a holistic basis, a larger and more natural corpus of language material than that represented by individual-element test questions.

Finally, the integrative-sociolinguistics period extends from the late 1960s to the present. This period focuses on 'communication' as functional-oriented or proficiency-based language teaching and learning in reaction to the dissatisfaction with structural linguistic theory. That is, the use of language for real-life communication came to be regarded as of great importance, and it provides us with a turning point from discrete point to an integrative approach. These assessment procedures to test more global than one thing at a time have come to be popular and a hot issue since the 1970s. The typical examples of integrative tests are cloze procedure, dictation, oral interview, easy writing, and so forth.

During these periods, discrete-point tests and integrative tests contested for primacy. But Carroll (1980: 12) maintains that both discrete point and integrative tests fall short of the ultimate goal:

Conventional tests ignore certain features of genuine language use, which are that language is authentic, contextualized, based in performance and assessed in

terms of behavioral outcomes. Indeed, it could be claimed that, at present, we are not testing genuine communicative performance at all, but an artificial, language-like behavior.

Hand in hand with the new conceptions in language theory was the continued development of English for Specific Purposes (ESP). Moreover, the development of detailed models or taxonomies of skills (Munby, 1978; Van Ek, 1976), reflected a new concern for functional, needs-related criteria, rather than the traditional, linguistic-based grid. They represented a fundamental shift in the focus of language teaching, with traditional language skills following from an identification of need.

2.2. TOEFL

The TOEFL exam developed by ETS in 1965 originally consisted of five sections. It was later reduced to three, due to some overlap, in 1976. Currently it consists of three sections: a listening comprehension section, a structure and written expression section, and a vocabulary and reading comprehension section. All sections are made up of discrete point, multiple-choice test items, allowing for highly reliable calculations to be computed and machine scorability. The statistical reliability of the TOEFL exam has been repeatedly confirmed, and there has been continued development of more sophisticated psychometric procedures. However, the overwhelming concern with statistical reliability, coupled with changes in linguistic and language learning theory since the development of the TOEFL, have led

many researchers to question the validity of the test. Oller and Spolsky (1979: 96) raise several of the major issues in a review of the TOEFL:

If a test were developed strictly on the basis of psychometric procedure without calling into play what is known of language competence from a linguistic point of view, it is just possible that it could be increasingly biased as items were constructed more and more to confirm statistically to the unique characteristics of a somewhat unusual population of speakers.

In terms of 'language competence,' research since the development of the TOEFL has shown that considerable doubt can be thrown on the claim that testing by an individual item and one skill gives a valid measurement of language proficiency (Oller, 1972, 1979; Bachman, 1982, 1985). Low correlations between the subsections of the TOEFL exam indicated to the developers that they were indeed measuring different components of language ability. Paradoxically, it is precisely these correlations that indicate a lack of validity to proponents of integrative testing. Maintenance of the statistical reliability between subsections of the exam is achieved in part by norm-referencing, i.e. standardizing candidates' results against each other rather than against any outside criteria. Oller and Spolsky (1979) contend that by standardizing results against a non-native population, the TOEFL exam is in danger of moving continually further away from a valid measure of how non-native performance will compare to the performance of native speakers. This is particularly significant for the TOEFL, as the

majority of participants take the exam with the purpose of admission to higher education, where they will be pitted against native speakers.

The TOEFL exam exemplifies the indirect test method, i.e., inferring one kind of performance from the measurement of another related performance (Henning, 1987). Henning describes test types as a continuum on which lies 'natural-situational' at one end and 'unnatural-contrived' at the other. Henning cites the oral interview as an example of the former, (a direct tap of the ability in question), and multiple-choice recognition as an example of the latter. Arguments against indirect testing rest on the assumption that indirect testing measures language proficiency skills less adequately than direct methods. The exact position here is unclear, and currently under debate; however, studies have shown that indirect testing can have a negative effect on examinees and a confounding effect on results.

2.3 IELTS

The IELTS, which superseded the earlier English Language Testing Service (ELTS) introduced in 1980, is jointly managed by the British Council, the University of Cambridge Local Examinations Syndicate (UCLES) and the International Development Program of Australian Universities and Colleges (IDP). It is used world-wide as one of the most remarkable standardized tests, which assesses the English proficiency of non-native speakers of English who intend to study in English-medium educational institutions. That is, the IELTS purports to provide a profile of the language performance of

each candidate which can be compared with that required for the course of study which he/she wishes to pursue (Davies and West, 1989).

The IELTS is quite different from the TOEFL in that the former measures the four skills of English - listening, speaking, reading, and writing - and that it does not maintain only the multiple-choice items. One typical characteristic of the IELTS is that the sub-tests of listening and speaking contain the same items which the examinees will solve, while those of reading and writing have four different modules, depending on the course study the examinees desire to take. That is, these four modules measure reading and writing in an academic or general training context. It is assumed that the main reason for selecting one of the four modules is due to the program of English for Specific Purposes, which is important for the efficiency of language.

The format of the IELTS maintains a minimum of MC item-types. For example, in a test of listening comprehension, test item-types based on the survival topics related to travel, accommodation, health, welfare, education, etc. are given with MC, form filling, gap-filling, True-False, short-answer questions. Especially, it is pointed out to the IELTS's credit that extensive use of MC questions is avoided in this test so that guessing does not become an important factor. Especially, in the test of speaking and writing section, these sub-tests of the IELTS directly measure the writing and speaking ability of the examinees, though they are very similar to the Test of Spoken English (TSE) and Test of Written English (TWE) of the TOEFL. As a result, the IELTS is an example of a good testing

system because it directly measures the language skills based on both approaches (discrete point and integrative) in language testing.

3. English Language Testing in Korea

3.1 Development of the English Tests for College Entrance

Since the introduction of English education in Korea, trends and methods of assessment of English have undergone a lot of changes in their theories, techniques, and formats. Regardless of whether the assessment of English is an achievement examination or a proficiency test, all three developmental stages suggested by Spolsky (1978) were witnessed in Korea. Translation from English into Korean and very rarely from Korean into English was once the sole technique and format of assessment. However, with the arrival of new generation testers during the 1970s, discrete test items were introduced and psychometric analysis techniques were employed to interpret test results. During the 1980s, a variety of tests such as problem-solving tests, pragmatic tests, maze tests, cloze tests, lecturette tests, etc., were introduced, though only a few were accepted for their practical value. To add global value to the discrete items, testers began a few years before to create contexts for their discrete items by providing longer passages for the items of reading comprehension and listening comprehension.

The English assessment methods which drew national attention

were selection tests, especially those which were required for college entrance. As the Ministry of Education institutionalized a compulsory test for prospective applicants for undergraduate colleges, namely, the Preliminary Test of the College Entrance (PTCE) in the late 1960s, English became a part of the test. Up until the beginning of this decade, the English tests consisted mainly of items of grammatical structure and reading comprehension. Structural items were on grammatical points which would cause confusion on the part of the learners. Reading comprehension was a sort of multi-purpose set of varied types of questions, such as sentence completion, English-Korean translation, rephrasing, comprehension of the content, inversion, etc. Other minor test items were of spelling, punctuation, pronunciation by phonetic alphabet, vocabulary, etc. Especially, vocabulary items included synonyms, antonyms, word definitions, inflections, affixes, idioms, and so forth.

At the beginning of the 1990s, the PTCE was renamed 'the College Scholastic Ability Test' (CSAT) for college level studies. At the start of the CSAT system, an important change was made in the English test format: that is, items of listening comprehension were included for the first time ever. The seven experimental tests, including listening comprehension test items were conducted during the two year period from 1990 to 1992. Eventually, the CSAT, a new SAT with an English listening comprehension test, was given in July, 1993. Though the introduction of listening test format did not make a complete improvement of the SAT, it encouraged teachers, students, parents and other concerned parties to pay considerable attention to spoken English as well as written

English. As a matter of fact, the washback effect of the CSAT English test has been absolutely strong in both public and private sectors, and English has been a subject for which examinees and their supporters willingly sacrifice their time and effort. With the listening comprehension test, the examinees are now preparing themselves for tests of spoken English as well as written English. At last, the CSAT English test has brought an innovation of English education to Korea.

3.2 English Test in the SAT

The SAT had continued to be administered until 1992, taking over in 1969 from the PTCE, which was the changed form with 30% subjective items in addition to the conventional MC items. It was made up of 42 MC items (42 points) and 8 subjective items (18 points). These items can be classified in the following five categories:

The first category was that of pronunciation and stress, which was for measuring listening and speaking ability indirectly. The number of the items of this category included in the SAT was seven in 1988, one in 1989, two in 1990, one in 1991, three in 1992, respectively. The drastic difference in the the number of the items in 1988 SAT (seven items) and 1989 or 1991 SAT (one item) is clear evidence of the inconsistency in the test item allocation. Moreover, it should be pointed out that testing spoken language ability with a written test type was not desirable in any way. It's because in the written test type, the student could give correct answers by mere memory without any adequate listening or speaking ability.

The second category was that of every day English expressions. As in the category of pronunciation and stress, this category contained useful expressions frequently used every day in places like restaurants, streets, banks, stores, or airports. Six items in this category were included in 1988 SAT; three items in 1989 SAT; two items in 1990, 1991, and 1992 SAT. The number of the items was too small, considering that speaking ability is one of four essential skills.

The third category included vocabulary, spelling, synonyms, antonyms and idiomatic expressions. A word may have a number of meanings in the sentences, so it should be tested in the context. In this respect, testing spelling, synonyms and antonyms was not so advisable, because they would only measure the examinees' mechanical memorizing ability. Vocabulary test items should be used for testing his or her inferential ability about their meaning in the context because vocabulary ability is closely related to reading comprehension ability.

The fourth category was that of grammar and composition. Accuracy was most important in this category. Unfortunately, however, the SAT tried to test grammatical knowledge or grammar itself by rote memory or through complicated knowledge of grammatical usages. Composition tests took the guided composition forms asking examinees to translate Korean into English. Because of possible difficulties of scoring, these tests were structured so as to make the examinees answer with basic grammatical knowledge and idioms. This category could not adequately test the examinees' linguistic ability.

The fifth category was that of reading comprehension. Reading comprehension was the biggest part in the test and also a

category most examinees found difficult. Reading comprehension comprised realistic comprehension, inferential comprehension, and general comprehension, including judgement, appreciation, and so on. The reading comprehension test measured the examinees' general ability to grasp the topic, main idea, and contents of comparatively long passages. It was desirable that much more items should be included to make the examinees try to understand the passage's logical characteristics and infer the main idea or topic and grasp the conclusion as in the current CSAT.

In summary, the SAT was principally a test for measuring the examinees' reading ability among the four skills of language, and it was also a test of linguistic knowledge, with its focus on linguistic usage. Therefore it could not measure the examinees' communicative competence properly.

4. The Analysis of the Foreign Language (English) Section of the CSAT

4.1 Objectives of Evaluation

To settle the situations of high school English education, the CSAT model was presented by the National Board of Educational Evaluation, and pre-tests for the experimental studies were performed seven times from December, 1990 to November, 1992. Since then, the actual CSAT has been performed four times.

This test evaluates not only the college or university

scholastic ability but also the degree of the achievement of high school English education objectives. Moreover, in Korea, it includes listening comprehension test, which was first introduced as the test form for spoken language. Thus, it has played an important role in settling the situation of English education in Korea.

While the SAT had evaluated the knowledge of English, the CSAT has especially attempted to evaluate the ability to use English for communication. Hence, it assesses the abilities to listen and comprehend English directly. Also it assesses the abilities to speak and write English indirectly. Considering all these situations, the objectives of the foreign language(English) of the CSAT are as follows (Woo, et al., 1993:105) :

1. To assess the English use ability.
2. To assess reading comprehension ability directly rather than accuracy.
3. To assess listening and reading comprehension ability directly.
4. To assess speaking and writing ability indirectly.
5. To focus on meaning- comprehension.
6. To help the settlement of English education.

The objective of the CSAT is not to evaluate the knowledge of English usage but to evaluate the fluency of listening, speaking, reading, and writing in English. The items of this test are made of vocabulary, grammatical judgement, realistic comprehension, inferential comprehension, and application related.

4.2 Contents of Test Items

Kim (1994:30) presented the differences of the CSAT from the SAT:

1. To Test language use ability rather than the accuracy of language structure.
2. Not to test pronunciation and stress by paper-and-pencil tests.
3. Not to include spelling or simple vocabulary tests.
4. Not to include the items without contexts.
5. Not to include the items using grammatical words.
6. To include the items asking socio-linguistic appropriateness.
7. To include more thinking ability-based items than simple memorization-based ones. To test grammatical judgment, but not to present sentence-level but paragraph passage-level items indirectly.
8. To test speaking and writing ability indirectly by presenting discourse or paragraph passages.
9. To emphasize communicative ability and include listening comprehension test.
10. To emphasize communicative ability and to test reading comprehension and language use ability.

Considering that the SAT and college-managed test increased emphasis on reading ability, the CSAT made real school education change its teaching-learning method to emphasize both spoken and written English. The CSAT assesses the four skills of listening, speaking, reading, and writing as shown in the following table:

Behavioral Contents	Vocabu- lary	Gram- matical Judg- ment	Comprehension			Applica- tion	Total	Ratio (%)	
			Realistic	Inferen- tial	General				
Listen- ing	Conversa- tion		3	3	1		7	14	
	Discourse				1		1	2	
Speak- ing	Conversa- tion			1		2	3	6	
	Discourse			1		1	2	4	
Writ- ing	Sentence						0	0	
	Paragrap h		2			3	5	10	
Read- ing	Editorial Sentence	2	2	3	3	3	13	26	
	Explana- tion Sen.		1	2	6	1	10	20	
	Literatur e			1	1	1	1	4	8
	Practical Sentence					2	2	4	8
	Others	1						1	2
Total Numbers of Items		3	5	9	15	12	6	50	
Ratio (100%)		6	10	18	30	24	12		100

On the basis of this table, analysis of the CSAT language ability test may be summarized as the following:

First, eight listening comprehension items were presented among 50 items, so these listening items charged 16%, and vocabulary expressions were easier than reading contents. The items emphasized realistic comprehension of conversations and discourses and inferential comprehension. And they also included general comprehension ability items like judgment,

appreciation, and so on.

Second, five speaking test items were indirectly presented through conversations or discourses. They emphasized contextually appropriate expressions in concrete situations. Especially, to evaluate the communication appropriateness, the items asked whether the examinees knew how to restructure various sentences presented in conversation using various expressions.

Third, there were five writing test items (10%), including grammatical judgment, logical order or development, conjunction of paragraphs, which were all tested indirectly.

Last, reading test items covered more than about 60%, based on reading comprehension ability of college textbooks. These items attempt to evaluate general reading comprehension ability by asking examinees to read various long passages or articles and find topics, main ideas, realistic comprehension, reasoning comprehension, appreciation, and judgment.

The CSAT tests given in 1994 and 1995 were like 1993's test, but they vary in two ways. First, the ratio of the listening comprehension test has increased year by year, the 1996's listening test will be 17 items, including indirect speaking items, and it will cover 30% (24 points for total 80 points). Second, since 1994's test, the principles of scoring points of all items (0.6, 0.8, 1.2) will be continued, and furthermore, the points will be increased like 1.0, 1.5, 2.0 by difficulty degrees.

4.3 Orientation of Evaluation Reform

According to the "The Reform Ways of the CSAT in 1997 Year" held on the 28th of August, 1995, the CSAT system and test

contents will be changed in the following ways:

First, in all sections of the CSAT, items will be increased, so foreign language(English) test items will be increased to 55 items (50 → 55) and total 80 points (40 → 80). Especially, to increase the discrimination factor, different scoring points(1.0, 1.5, 2.0) will be applied by difficulty degrees for each item. The total test time will be same as in the past, 80 minutes, but the listening test will cover 17 minutes.

The CSAT, which will be performed in 1996, is going to be taken in the same frame as the four CSATs. The principle of different scoring points is based on not difficulty degrees but thinking ability. Therefore, items like short sentences or passages, simple thinking ability-solving vocabulary, and grammatical judgment items, should be allotted 1.0 scoring points, but higher level thinking ability-solving items, which demand general thinking and application abilities, should be allotted 2.0 scoring points.

Especially, considering the principles of high school curricula emphasizing listening and speaking ability, it is necessary to lengthen the conversations or discourses, to increase the quantity of information, and to decrease the repetition of cue contents of listening items. In the CSAT, speaking test, strictly speaking, is a part of the listening test. Absent of any speaking ability test models, the reformed test formats are not persuasive. The speaking ability test, which has been presented indirectly through written forms, is going to change. Written form-based reading and writing tests should be continued, but items can be presented through one 110-words passage (unlikely past 60-90 words). The more words, the longer

passages take to read. In other words, the reading and writing test items, which will be taken in 1996, will increase by two items over the 1995 test. On the contrary, testing time will decrease by four minutes (67 mins → 63 mins), so it is necessary to read faster. Generally speaking, considering that the language fluency test is not a power test but a speed test, presenting one or two long passage items (which have more than 100 words) is a more reasonable way to increase difficulty degrees and discrimination (Jeon, 1995:176).

Next, through the table of the foreign language (English) section contents and behavioral table and items, the various items which attempt to test English use abilities pointed out in the high school English curricula should be presented. This will help resettle high school English education and promote desirable English teaching-learning methods which can develop the students' communicative competence.

5. Conclusion

Overall, in Korea three developmental stages of language testing have been witnessed and developed at the same time within a 20-year period. One of the factors which accelerated the development of language testing is the nationally known importance of the college entrance examination. The 1993 advent of the CSAT, which introduced the first listening comprehension test in 1993, was influential in dictating secondary educational institution's teaching methods.

This paper reviewed not only the theories of language testing

but also two typical types of international standardized English proficiency tests, TOEFL and IELTS. It also reviewed and discussed two different kinds of English proficiency tests, such as the SAT and the CSAT administered by NBEE. Through a comparison of these tests, this paper analyzed the problems of the CSAT and suggested the desirable directions for its use in the Korean educational situation. The problems and the desirable directions are summarized as follows:

First, the concept of the CSAT is so vague that we should clearly define the objectives of assessment. For example, NBEE defined these objectives in the following way: "The test attempts to evaluate the intellectual thinking ability of the applicants which is needed to be successful while they are attending colleges and universities.... The test questions are formulated from the integrated subject matter of the high school curricula as it is taught in high schools." Accordingly, the level, the contents and methods of assessment should be clearly illustrated.

Secondly, the CSAT does not faithfully assess four skills of English due to the drawback of the MC test. Especially, it cannot directly measure the examinees' speaking and writing ability. Moreover, it is desirable to change the questions currently in Korean into English.

Thirdly, the behavioral factor of the CSAT is not empirically justified. Therefore, it is difficult to justify the allotment of scores per item. Much research should be carried out to justify the empirical and construct validity for a valid behavioral section.

Language testing in recent trends should be direct and

authentic based on the kind of integrated performance in real-life communication. Thus, gradual efforts are needed to improve the CSAT so that it can validly measure the four skills of English.

REFERENCES

- '97 대학수학능력시험 개선방안 연구회. 1995. '97 대학수학능력시험 개선방안.' 국립교육평가원.
- 김기화. 1994. '영어과 교수-학습 및 평가방법의 개선방향.' 전북중등교육 연구회지, 제4호, 11-17.
- 우중옥 외. 1993. 대학수학능력시험의 영역별 출제모형 정립 및 모형활용 방안연구. 국립교육 평가원.
- 심재기 외. 1994. '대학수학능력시험 문제의 질적 향상을 위한 출제 방향 및 방법의 개선 연구.' 1995 학년도 대학수학능력시험 출제위원회.
- 전병만. 1985. 언어능력 평가. 서울: 한신문화사.
- _____. 1988. 'New Directions of Language Testing.' 영어영문학 제 11집. 한국영어영문학회 전북지부, 237-272.
- _____. 1991. 영어시험의 분석과 출제경향. 전북대신문.
- _____. 1995. '96학년도 대학수학능력시험 외국어(영어)영역의 방향.' Foreign Languages Education, No 1. 한국외국어교육학회, 173-177.
- Bachman, L. F. 1982. 'The trait structure of cloze test scores.' *TESOL Quarterly* 16, 61-70.
- _____. 1985. 'Performance on cloze tests with fixed-ratio and deletions.' *TESOL Quarterly* 19, 3, 535-56.
- _____. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Canale, M. and M. Swain. 1980. 'Theoretical bases of communicative approaches to second language teaching and testing.' *Applied Linguistics* 1, 1-47.
- Carroll, John B. 1961. 'Fundamental consideration in testing for English proficiency of foreign students.' *Testing the*

- English Proficiency of Foreign Students*. Washington, D.C.: Center for Applied Linguistics, 310-48. Reprinted in Allen and Campell (1972), 313-20.
- _____. 1980. *Testing Communicative Performance*. Oxford: Pergamon.
- Clark, John L. D. 1983. 'Language testing: past and current status-direction for the future.' *The Modern Language Journal* 67/ iv, 431-43.
- Davies, Susan and Richards West. 1989. *English Language Examinations*. London: Longman.
- Educational Testing Service. 1970. *Manuals for Peace Corps Language Teachers*. Princeton, N.J.: ETS.
- Hadley, A. O. 1994. *Teaching Language in Context (2nd Ed.)*. Boston, Mass.: Heinle and Heinle.
- Harris, David P. 1969. *Testing English as a Second Language*. New York: McGraw-Hill.
- Henning, G. 1987. *A Guide to Language Testing*. Cambridge, Mass.: Newbury House.
- Hughes, Arthur. 1989. *Testing for Language Teachers*. Cambridge: Cambridge university Press.
- Hymes, D. H. 1972. 'On communication competence.' In Pride, J.B.J. and Holmes (eds.) 1972, 269-93.
- Lado, Robert. 1961. *Language Testing*. New York: McGraw-Hill.
- Liskin-Gasparro, Judith E. 1982. *Oral Proficiency Testing Manual*. Princeton, N.J.: Educational Testing Service.
- Munby, J. 1978. *Communicative Syllabus Design*. Cambridge: Cambridge University Press.
- NBEE. 1996. *Guidelines for the College Scholastic Ability Test*. Seoul: National Board of Educational Evaluation.

- Oller, J. W. 1972. 'Scoring methods and difficulty levels for cloze tests of proficiency in ESL.' *Modern Language Journal* 56, 151-58.
- _____. 1976. 'A program for language testing research.' in H. Douglas Brown (ed.) *Papers in second language acquisition: Proceedings of the sixth annual conference on applied linguistics at the University of Michigan. Language Learning Special Issue Number 4*, 141-65.
- _____. 1978. 'Pragmatics and Language Testing.' In Bernard Spolsky (ed.) *Papers in Applied Linguistics: Approaches to Language Testing*. Arlington, Virginia: Center for Applied Linguistics.
- _____. 1979. *Language Tests at School*. London: Longman.
- Oller, J. W. and B. Spolsky. 1979. 'The test of English as a Foreign Language' in B. Spolsky (ed.): *Advances in Language Testing*. Alington, VA: Center for Applied Linguistics.
- Spolsky, B. (ed.). 1978. *Approaches in Language Testing, Advances in Language Testing, Series: 2*, Arlington, Va.: Center for Applied Linguistics.
- Savignon, S. J. 1972. Teaching for a Communicative Competence: a research report. *Audiovisual Language Journal* 10(3), 153-62.
- _____. 1983. *Communicative Competence: Theory and Practice*. Reading, Mass.: Addison-Wesley.
- _____. 1985. 'Evaluation of communicative competence: the ACTFL proficiency guidelines,' *Modern Language Journal*, 69, 129-33.

- Upshur, John A. 1962. 'Language proficiency testing and the contrastive analysis dilemma.' *Language Learning* 12, 132-37.
- Van Ek, J. A. 1976. *The Threshold Level for Modern Language Teaching in Schools*. London: Longman.
- Weir, C. J. 1990. *Communicative Language Testing*. Prentice Hall International (UK) Ltd.
- Widdowson, H. G. 1978. *Teaching Language as Communication*. Oxford: Oxford University Press.

Abstract

English Language Testing in Korea

- With Special Respect to the College Scholastic Ability Test

Jeon, Byoung-Man

This paper briefly reviews the theories of language testing and English language tests themselves as 'backwash', which can be harmful or beneficial to language teaching/learning methods in Korea. Based on the discrete-point approach, the College Scholastic Ability Test (CSAT) by the National Board of Educational Evaluation (NBEE) employs an objective test format of multiple-choice items to secure the reliability of the test scores. But one crucial drawback of the CSAT is that it does not directly measure the functional proficiency, and therefore, this test may lose content validity. That is, the CSAT is not a direct and authentic test. In spite of this weakness of the CSAT, this paper also points out the assets of the test, namely, its high reliability and scientific accuracy, which are crucial in a test-dependent society like Korea. Also, it is helpful to compare the CSAT with two typical types of international standardized English proficiency tests, Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS). Through the comparison of these tests and analysis of the CSAT, this paper discusses the problems of English tests used in Korea and suggests the desirable directions for appropriately using the CSAT in the Korean educational situation.