

THE PLACE OF GENERAL ACHIEVEMENT TESTING IN THE FRESHMAN ENGLISH LANGUAGE PROGRAM

メタデータ	言語: English 出版者: 明治大学経営学部人文科学研究室 公開日: 2012-05-16 キーワード (Ja): キーワード (En): 作成者: Vance, E.Johnson メールアドレス: 所属:
URL	http://hdl.handle.net/10291/11979

THE PLACE OF GENERAL ACHIEVEMENT TESTING IN THE FRESHMAN ENGLISH LANGUAGE PROGRAM

Vance E. Johnson

CONTENTS

1. Introduction
2. Discussion of Present Situation
3. Testing Considerations and Selection
4. Selection of Criterion Measures
5. Scoring the Test
6. Pre-test and Post-test Evaluation
7. Some Possible Applications of Curriculum Variables
8. Some Suggestions on Test Construction
9. Conclusion
10. Bibliography

1. Introduction

The aim of this short article is to show how the introduction of a testing program at the freshman level will greatly enhance the total English language program for universities that do not at present provide such a system. It will be shown that the best type of test is a general achievement test that can be used for both prognostic and diagnostic purposes if the test is given as a pre-test to all incoming freshman and a post-test at the end of the year. The impact of such a testing program on the students, the teachers and the total curriculum will be discussed. Finally, suggestions will be given as to the design of such

a test and the best method of scoring.

2. Discussion of Present Situation

“The unexamined life is not worth living.” (Socrates: in Plato’s Apology)

Most Japanese universities and junior colleges demand that all freshman students take credit courses in the English language. Successful completion of these courses is necessary for advancement to higher-level English classes and for graduation. As these general English courses are compulsory for all students, and most schools recognize the need to limit the size of classes in order to provide a better atmosphere for more individualized language learning, universities spend a great deal of time and money on these programs. Since the numbers greatly exceed the capacity of the English staff of any individual university, much of the program is staffed with part-timers, usually professors from other colleges. Often native speakers are also hired part-time to help teach the large number of freshman classes. The result is a very large program that takes a great deal of investment in time, money and the use of often limited university facilities. The setting up, scheduling and implementation of the freshman English program represents a firm commitment on the part of the university to the necessity of the Japanese student to learn the English language.

For these reasons alone, the freshman English program should be constantly evaluated and reviewed, in order to make sure that all these resources are being used most effectively.

The problem is, however, that for most universities at present there is no way to evaluate whether the general English program is making maximum use of its potential, or indeed whether it is succeeding or not. The reason for this is that whereas the university provides the classrooms and teachers, it gives no guidelines regarding the aims and goals of the program, and provides no objective method for students and teachers to measure what progress has taken place. Because actual goals and instruction are left entirely up to the individual teacher, there is great diversity within the program. In regard to method, it ranges from teachers who prefer only translation to those who forego any text and conduct every class as “free discussion”. In regard to level it ranges from linguistics or literature professors teaching their speciality to teachers who, preferring remedial work, use current lower secondary textbooks in order to instill basic sentence patterns and vocabulary.

The genial anarchy that results is justified under the auspices of “academic freedom”,

and one reason why changes in the lower level university program have not been made up to now is that university teachers greatly fear any interference or controls in the classroom.

The problem then is to provide some sort of general aims and goals for the freshman English program and a means to evaluate whether the program is meeting those goals or not without infringing on the traditional academic freedom allowed in Japanese universities. Fortunately there is a way to reconcile these conflicting needs, and that is to implement a program of pre and post testing of all freshman. Such a test would by its very existence define the goals and criterion measures for the English program, for what is included and what is excluded would indicate what is considered important by the university authorities, and signal it to students and teachers alike. By having an objective measure of student progress, the school would have the option of experimenting with different variables to see if the program could be made more effective and more cost efficient. Psychologically, it would also indicate that the university is serious in its intention to teach English and for the students to learn English which could not but help build stronger incentives in both teachers and students. Such a test, given before classroom instruction has started and after it has ended, would in no way interfere with academic freedom in the classroom and so should meet with little opposition from teachers on that basis. Individual students would have a better way to assess their progress than the grade given by one or two professors by comparing their scores with the progress made by the total freshman class.

If the unexamined life is not worth living, a case could be made that “the unevaluated foreign language program is not worth having” or at the very least such a program allows waste and inefficiency to continue year after year with detrimental effect on university, teacher and student.

3. Testing Considerations and Selection

Assuming that the benefits of having some sort of testing system are recognized, the next problem is to select a type of test that will best evaluate the objectives of a university freshman English program. Testing specialists recognize four different kinds of tests for language purposes although they disagree as to their definition. They are prognostic, diagnostic, proficiency and achievement tests.⁽¹⁾

(1) Valette uses a different nomenclature, calling a prognostic test an “aptitude test”, making a distinction between what she calls achievement tests and “progress tests” and providing no classification for diagnostic tests. See Valette, 1977.

A prognostic test is given with the purpose of finding information that will tell you how a student will do in a future course of study. Although many people think of prognostic testing in terms of such foreign language aptitude tests as the MLAT (Modern Language Aptitude Test) (Carroll and Sapon, 1959), one can also think of prognostic testing in terms of the testing and placement of students who have previously studied the language.⁽²⁾ A general achievement test could therefore be used in a prognostic way if it is used to select students for streamed freshman English classes.

According to Mackey, the purpose of a diagnostic test is to find out what remains to be taught, for as he says :

“What the class has been, taught, however, is no indication of what it knows.”⁽³⁾

The function of the diagnostic test then is to provide information on the students' knowledge of the language and not to assign points or grades to the students.

Proficiency tests are given to measure the students' actual ability to use the language⁽⁴⁾ and so the test is not based on classroom learning but on how well the student functions in actual situations. Proficiency testing is thus tied up with communicative competence as opposed to language competence.⁽⁵⁾ One of the best known tests developed to measure communicative proficiency is the U. S. Department of State Foreign Service Institute test officially called the “Absolute Proficiency Rating” system, but perhaps better known as the “FSI Language Proficiency Interview”. It consists of an interview session to determine which of five levels the student falls into, ranging from elementary proficiency (Level 1) to native proficiency (Level 5).

Achievement tests are given to determine how much a student has mastered within a given language area. They include the traditional tests given at the end of a course to measure how much of the material a student has mastered. The meaning of achievement tests may be broadened, however, to test the student's mastery of course content on a more global basis.⁽⁶⁾ Such tests may be called “general achievement tests” and they differ from

(2) See Clark, 1972, for a detailed definition of prognosis and for his separation of prognostic testing into “selection” and “placement”.

(3) Mackey, 1965, p. 405.

(4) Clark calls it “real-life competency”. Clark, 1972, p. 5.

(5) See Politzer, 1972, pp. 138-140 or Clark, 1972, pp. 118-132 for a more detailed discussion of communicative vs. language competence.

(6) See Clark, 1972, p. 4.

proficiency tests in that the content of the test is kept “within the general framework of the student’s formal course work up to that time⁽⁷⁾”.

“Although meaningful part scores for certain sub-sections of the test can in some cases be produced, the more usual outcome of a general achievement test is a single total score which indicates only how well the student dealt with the content of the test *as a whole* and without regard to specific linguistic details of his performance. With appropriate statistical caution, students within a class can be ranked on the basis of total test score, and where normative information is available, the performance of individual students or of the class as a whole can be compared to that of larger groups of students who have taken the same test. Scores on general achievement tests can also be used for selection or placement purposes.”⁽⁸⁾

From the brief discussion above, it can be shown that on the basis of testing principles, a test of the general achievement type would be best suited to evaluate the effectiveness of the freshman English program and evaluate the progress of individual students within that program. As very few entering Japanese students have had contact with English outside the classroom situation, it would be unrealistic to give them “real-life” proficiency tests, especially since the end of secondary training in English in Japanese high school is to pass the entrance examinations for colleges and universities. A general achievement test could be validly used as a prognostic measure to select students for various courses within the curriculum, if the university so desires. A well-made general achievement test, with an emphasis on global rather than discrete-point testing⁽⁹⁾ could best test the broad range of teaching methods, levels, and materials found in the present typical freshman English program.

4. Selection of Criterion Measures

This article has been written to see forth the hypothesis that by instituting a testing program in the freshman English program, “better results” will accrue than having no method of course evaluation. If it is possible to get all the disparate viewpoints of faculty, students and administration to agree on any one thing, it is that all would agree that they want

(7) Ibid.

(8) Ibid., p. 5.

(9) See Valette, 1977, p.12 for information on discrete-point and global testing.

“better results”. Unless we think of “better results in a scientific way, however, the term is meaningless as it means so many things to so many different people.

“‘Better results’ implies the measurement of outcomes. It implies a criterion test. If the criterion test does not measure what the agreed upon outcome of instruction should be, experimentation is meaningless. Much of the perpetual discussion as to what method or procedure in foreign-language teaching is ‘better’ is confused and confusing, because disagreement as to methodology is accompanied by disagreement as to what the objective of instruction should be and how it can be measured.”⁽¹⁰⁾

Nowhere does the above passage better apply than in the various language departments of Japanese universities. What is implied by “better results” in English language teaching in Japan has had effects similar to the American Civil War, it has set teacher against teacher and friend against friend, and in the case of some large and prestigious universities, has divided faculties to the point that dialogue no longer takes place between the opposing factions.

I don’t wish to digress from my point, but it is important to emphasize that the setting up of criterion measures has proved a formidable and often impossible task in Japanese universities.

What needs to be emphasized then is that a general achievement test presupposes no specific method or objective of instruction. Therefore, the implementation of such a test does not necessarily favor one group over another. It does, however, open the way to dialogue concerning what should be tested. If a test is judged worthwhile, some area of knowledge must be tested in a certain way and decisions must, willy-nilly, be made.

If the faculty of a university decides that the objective of freshman English should be to hone the translation skills of students so they are better prepared for their upper-level classes, then let that be the criterion measure. At least it will be clear to the students what is expected of them and they can study accordingly. At the other extreme, if communicative speaking ability is the main purpose of freshman English, then let the general achievement test focus on that. The point is that the program needs to have SOME goal or goals that are made clear to teachers and students alike, so that the program can move in the direction of the decided goals and an evaluation can be made of the effectiveness of the program.

⁽¹⁰⁾ Politzer, 1972, p.135.

Most schools would probably find themselves in the middle, with a wish for students to gain the skills needed for further study at the university and also to gain some ability to “use” the language and apply their study to “real-life” situations. For such faculties, they may wonder how to make a test that would actually test such a diversity of purposes.

Later in this paper I will talk about some possible solutions; but first let us look at some models of possible classification of language goals and objectives.

In regard to the taxonomy of language learning, two models, the Valette-Disick Taxonomy and The Rivers Model are helpful, and they are reproduced here as an aid for those readers who do not have access to the books or the inclination to look them up.⁽¹⁾

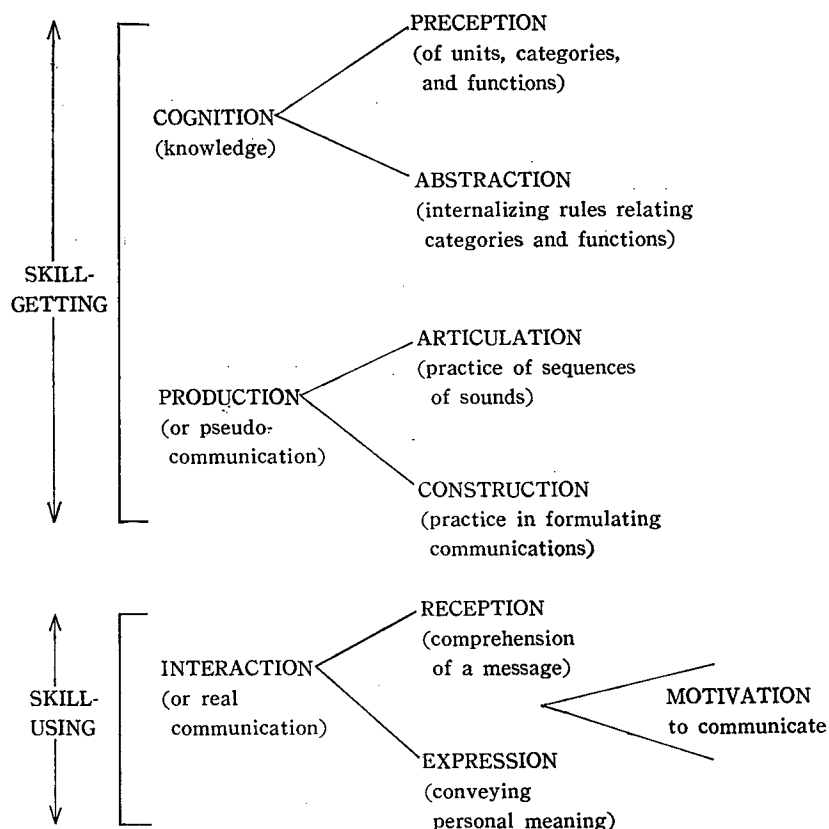
Valette-Disick Taxonomy
Figure 1 SUMMARY OF THE SUBJECT-MATTER TAXONOMY

Stage	Internal Behavior	External Behavior
1. Mechanical Skills : The student performs via rote memory, rather than by understanding.	<i>Perception :</i> The student perceives differences between two or more sounds or letters or gestures and makes distinctions between them.	<i>Reproduction :</i> The student imitates foreign-language speech, writing, gestures, songs, and proverbs.
2. Knowledge : The student demonstrates knowledge of facts, rules, and data related to foreign-language learning.	<i>Recognition :</i> The student shows he recognizes facts he has learned by answering true-false and multiple-choice questions.	<i>Recall :</i> The student demonstrates he remembers the information taught by answering fill-in or short-answer questions.
3. Transfer : The student uses his knowledge in new situations.	<i>Reception :</i> The student understands recombined oral or written passages or quotations not encountered previously.	<i>Application :</i> The student speaks or writes in a guided drill situation or participates in cultural simulations.
4. Communication : The student uses the foreign language and culture as natural vehicles for communication.	<i>Comprehension :</i> The student understands a foreign-language message or a cultural signal containing unfamiliar material in an unfamiliar situation.	<i>Self-Expression :</i> The student uses the foreign language to express his personal thoughts orally or in writing. He uses gestures as part of his expression.
5. Criticism ; The student analyzes or evaluates the foreign language or carries out original research.	<i>Analysis :</i> The student breaks down language or a literary passage to its essential elements of style, tone, theme, and so forth. <i>Evaluation :</i> The student evaluates and judges the appropriateness and effectiveness of a language sample or literary passage.	<i>Synthesis :</i> The student carries out original research or individual study or creates a plan for such a project.

(1) The charts are taken from Valette, 1977, Which were in turn taken from Valette and Disick, *Mod. Lang. Perf. Obs.*, page 41 and Wilga Rivers, *A Practical Guide to the Teaching of French*, pp. 3-5.

Rivers Model

Figure 2 PROCESSES INVOLVED IN LEARNING TO COMMUNICATE



It should be noted that in the Valette-Disick Taxonomy a distinction is made between internal behavior and external behavior, which Valette says is primarily the difference between listening and reading on the one hand and speaking and writing on the other. The Rivers model makes a distinction between "skill-getting" and "skill-using". Rivers' "skill-getting" corresponds to Stage 1-3 of the Valette-Disick Taxonomy and "skill-using" corresponds to Stage 4, the result of both models being "real" communication.

Testing specialists have traditionally broken down the skills to be tested into four types, listening comprehension, speaking, reading comprehension, and writing, and treated each one as a separate unit to be tested. This is done in order to provide a better measurement of the results. The following chart, taken from Clark, 1972, provides a good model of tra-

(12) Valette, 1977, p. 21.

ditional testing approaches.⁽⁴³⁾

Figure 3 Stimulus-Response Modalities for Testing Specific Skills^o

LISTENING COMPREHENSION

Stimulus: Spoken FL

Response: Spoken FL

Written FL

Written NL

Multiple-Choice†

SPEAKING

Stimulus: Spoken FL

Written FL

Written NL

Pictorial

Response: Spoken FL

READING COPREHENSION

Stimulus: Written FL

Response: Spoken FL

Written FL

Written NL

Multiple-Choice†

WRITING

Stimulus: Spoken FL

Written FL

Written NL

Pictorial

Response: Written FL

^o Does not include "passive" testing of speaking and writing

† Including "FL reading," "NL reading," "pictorial," and "pure" sub-categories.

As almost all research in foreign-language testing has followed a variation of the above model, teachers setting up a testing program would be well advised to stay within the model, so they may benefit from the many studies and statistical results of foreign-language testing using this model.

It would be up to the various English departments of the various universities to decide what weight to put on the skills of listening, speaking, reading, and writing, depending on how they select the objectives of their program. Possibilities can range from a test that equally tests all four skills to one that tests only reading comprehension or listening comprehension.

Since the scoring of a pure speaking or writing test requires much more time and man-

(43) Clark, 1972, p. 32.

power and is subjective rather than objective, I doubt that many schools will take up those options, though from the viewpoint of the goals of language learning being communication and ability in self-expression, such tests, if done, would truly become models for all schools in Japan to emulate.

5. Scoring the Test

Assuming that the hurdle of selecting criterion measures has been overcome and the areas to be tested have been decided, what is the best way to score the test? Considering the numbers involved (students, teachers, scorers, etc.) the easiest method of scoring that has reliability and validity⁽⁴⁾ should be considered. From my personal experience in constructing and scoring tests, I have found the multiple-choice format to be the best. This also is the opinion of testing experts.

“Multiple-choice format meets both of these scoring criteria in an exemplary manner, with the aid of a simple scoring key or stencil, the teacher can very quickly and easily score a multiple-choice test or, where this possibility is available, assign the task to an electronic scoring device. Scoring reliability of multiple-choice tests is virtually absolute: barring clerical errors, a given student performance (i. e., a particular answer sheet) can be evaluated again and again with precisely similar results.”⁽⁵⁾

Schools with smaller numbers of students and a more creative approach to language testing should certainly consider optional methods of scoring, and other methods available can be found in any of the testing books cited in this report, but it is assumed that most schools will take up the multiple-choice format as their choice of scoring a general achievement test given to all freshman.

6. Pre-test and Post-test Evaluation

Once the goals of the program have been decided and a test has been prepared that adequately tests those goals, and the test provides reliable scoring, we have the essential ingredients for evaluating the effectiveness of the program. This is done by giving the test as a pre-test to the incoming freshmen students before or as soon as classes start and again at the end of the school year. By comparing the pre-test and post-test scores and

(4) See Valette, 1977, Chapter 4, for details on testing reliability and validity.

(5) Clark, 1972, p. 40.

doing the required statistical analyses on them, a method of measuring total as well as individual progress is made possible. By changing variables within the program, the test provides a means of evaluating whether such changes have benefitted student learning or not. Once the basic mechanics have been set up and the test has been implemented, other educational uses of the pre-test and post-test are possible, if the university so wishes.

The pre-test could be used as a placement test, and students could be streamed into leveled classes, or special classes set up for the needs of special students. Many universities in Japan do not make provision in their English language program for either advanced students, such as those who have studied abroad and are therefore at a much higher level than the average freshman student, or for those students who are quite weak in basic English skills and would benefit from a special remedial class. The pre-test would be able to target such individuals and special classes with smaller numbers of students could be set up for them. Some Japanese professors equate such streaming to "elitism" and as opposed to the egalitarian nature of post-war Japanese education. This is nonsense. It would be similar to a music school putting piano students with the career and technical skills to play Beethoven sonatas together with those students still learning basic skills and to whom Burgmuller seems difficult.

The post-test would provide the student with an objective basis to evaluate his individual progress during the year and to compare his progress with that of his peers. Freshman grades only show what a student has learned in a particular class and give him no basis in deciding how his general language abilities have improved.

The teacher, if he so wished, could use the results of the post-test to evaluate his students' progress compared to the average progress of the total freshman class and use information to help plan changes in his next year's teaching. From personal experience, however, I have discovered that since individual teachers meet the average student only once a week, the input of a single teacher has little influence on the total average.

Of course personal contact with a superior teacher can greatly stimulate certain students to study much more, but the total learning average seems quite consistent year by year if the program is the same, and teachers come and go. It is this consistency that would allow a statistical evaluation of curriculum changes to see whether they benefit the program or

(10) I am basing this statement on several years work and results computed at the Foreign Language Center of my previous university.

not. If the university so wished, a certain level score of the post-test could be considered a minimum standard for (1) passing the freshman English program even though a student were to pass individual English classes or (2) admittance to upper-level English courses offered by the university. Individual teachers could also, for example, demand a certain level score for admittance to their classes, thus obviating the need for teachers to make up their own tests and score them themselves in order to determine which students to select for class in those cases where numbers of students are in excess of space available.

7. Some Possible Applications of Curriculum Variables

Assuming that a valid testing program has been set up, what are some of the variables that can be changed to improve the program? As this comes under the heading of research in Applied Linguistics, it would be useful at this point to get an overview of variables that can affect language learning. One of the most concise and useful explanations I have found is the following, taken from Politzer.⁽¹⁷⁾

Figure 4 A TYPOLOGY OF RESEARCH IN FOREIGN-LANGUAGE TEACHING

A. Teaching variables:

I. Teacher variables.

- age,
- sex,
- teaching experience,
- proficiency in the foreign language,
- knowledge of linguistics,
- years of residence in the foreign country,
- attitude toward :
 - (a) the student,
 - (b) the foreign language and culture,
 - (c) the native language and culture of the student.

II. Equipment variables.

- type of laboratory (audio-active, audio-passive),
- audio cassettes,

(17) Politzer, 1972, pp.141-142.

video tape.

III. Method variables.

(a) general :

“grammar-translation method”,

“audiolingual method”,

“direct method”.

(b) specific variables :

type and length of drills,

place of explanation,

manner of supplying meaning,

sequence of presentation.

(c) space and time variables :

number of contact hours,

spacing of contact,

class size,

time of meeting of classes.

B. General environmental variables :

Sociological environment in which instruction occurs,

Parental attitude toward language.

C. Student variables :

age,

sex,

knowledge of other languages,

previous language training,

motivation, interest,

aptitude for language study,

attitudes toward :

(a) foreign language and culture,

(b) teacher,

(c) method,

modality preference.

D. Language variables :

Type of language problem to which criterion measure is applied.

Complexity of language problem.

Nature of “interference” from L₁.

E. Criterion measures:⁽⁹⁾

- (a) language levels⁽⁹⁾
 - phonology-pronunciation,
 - morphology,
 - lexicon-vocabulary,
 - syntax.
- (b) language skills
 - listening,
 - speaking,
 - reading,
 - writing.

Let us take one or two examples to show how the testing program can help evaluate changing variables in the language program. Suppose that the university has built a new language laboratory and wishes to know its effect on listening comprehension. Selected classes could be given one or two periods of L.L. work and at the end of the year, listening comprehension scores could be compared with students who follow the old program without L.L. work. On the basis of such work the school could decide to what extent to integrate the L.L. into the regular curriculum.

In another case, suppose a school were contemplating a structured program in which two or three teachers used the same textbook and complimented each other's teaching (for example one teacher working on translation and grammar, another on the knowledge and information given in the passage and another, discussion and writing related to the passage). A few trial classes could be set up and on the basis of the results of the testing program a decision could be made whether expanding the program would be worth the time and effort.

In another case, suppose the school were thinking of expanding the required hours for

⁽⁹⁾ Although this category is given in Politzer, the sub-variables in this paper are mine.

⁽¹⁰⁾ This breakdown of language testing into levels and skills and their intersection is described in Carroll, 1961, and also in Cooper, 1967.

language training (or perhaps the opposite). What effect would this have on the language ability of the students? Without an adequate testing program only subjective opinions could be given and there would be no valid way to objectively decide the value of changing such an important variable as the number of contact hours.

Would a special intensive course for remedial students given during the first term bring them up to the level of the average student? Without a general testing program in place no one would ever know, for there would be no objective way to judge the level of the average freshman student of the university. In these and in countless other ways an established testing program at the freshman level would give the necessary answers for teachers and administrators to judge the worth of changing and upgrading the language programs of the university.

Having such a testing program would also stimulate individual research on the part of faculty members, which would lead to more understanding and to better motivation to improve their teaching.

8. Some Suggestions on Test Construction

Supposing that a faculty were to decide that some sort of testing program would be a valuable addition to their English language program. How should they go about it and what should they include? If several tests were to be made, how could standardization of difficulty be optimized? What criterion measures should be selected and what weight placed on them? How many items are necessary to measure the necessary ability and how long should the test be? These and other pertinent questions need to be answered at the outset.

It is not within the scope of this paper to answer in detail questions related to the actual construction of a general achievement test to be used in testing, but I feel some suggestions would be useful for those teachers who have had little contact with the actual construction of such tests.

Many people wonder if any one skill is predictive of the others to the extent that it can be used as a global predictor of language ability. At this point in time, the research results are inconclusive but my “intuitive” feeling is that listening comprehension may be the best indicator of the four skills. As most freshman English programs emphasize reading,

② This also ties in with the current feelings about the primacy of hearing in the learning process. See Postovsky, 1974, 1975.

however, in the absence of reliable data, it would be difficult to support the hypothesis that a listening test would adequately measure reading progress at this point in time. Most commercial general achievement-proficiency tests have listening comprehension combined, with either all or a subset of reading, vocabulary and structure. Commonly used tests are CELT (Comprehensive English Language Test), MELAB (Michigan English Language Assessment Battery) and TOEFL (Test of English as a Foreign Language).

My own feeling is that a university implementing a testing program need not try to imitate such ambitious tests. It would be better to select priorities and implement different skills at different times. This would allow maximum concentration on one level at a time and would guarantee a better final product that would fit the needs of the university and would not have to be tinkered with or perhaps even thrown out at a later date because of inadequate preparation. The university could either start with a reading or listening test. Because of the extra problems with the printing and handling of reading tests I recommend starting out with a basic listening test the first year and adding either vocabulary or reading the next year.

There are many leveled materials from which teachers may select and adapt for listening comprehension and reading tests, and many different kinds of questions can be handled by the multiple-choice format, making scoring either by computer or mechanical scoring machine a simple matter. I have found that it is very difficult to standardize vocabulary and syntax tests because it is often difficult to decide difficulty level and the vocabulary or syntax section of the test may end up being either more or less difficult than the previous year's test.

For the reading part of the test, as many universities give mostly reading-based entrance tests, giving a previous year's English entrance examination as a post-test and keeping a special record of the scores of the incoming freshman students on the English section of this year's test would provide one easier way to measure reading progress during the year, as most schools are very careful to standardize the tests from year to year. It would also be a good way to get more use out of old tests, as only those teachers who prepare them know how much blood, sweat and tears go into constructing the one-time entrance exam.

It is very important that test guidelines be drawn up so that people who were not part of the original planning are able to construct a test of the same type and level of difficulty. Also after four or five tests have been made, parts may be interchanged so that the whole

test doesn't need to be redone every year. After the test is given an item analysis should be made—from the very first year— so questions can be rated according to difficulty level and discrimination level and errors can be corrected before they become an integral part of the testing program.

In regard to more specific questions related to the construction of test items, books on test construction should be consulted. Many of the books listed in the bibliography will answer such questions in detail.

9. Conclusion

In this paper I set out to show that universities that do not now utilize general testing as a method to improve their freshman English program would discover many benefits from instituting such a program. Some of the benefits that are talked about in this paper include the following :

- 1) It provides a measure to evaluate yearly progress of the freshman class as a whole and also as individuals.
- 2) It makes clear the goals of the program to both teachers and students.
- 3) It shows in an objective way that the university is serious in its desire to have the students learn English.
- 4) It allows the curriculum to be experimented with to see if changes would produce better learning.
- 5) It could be used to set up special classes for students who would find little benefit in the normal curriculum.
- 6) It could be used as a means to admit students to upper level classes.
- 7) It could be used as a means to determine which students meet the university's minimum language requirements.
- 8) It could provide a basis for individual faculty research on language learning problems.

All or a subset of the above are possible according to the choice of the university. There has been much talk about the need for Japanese universities to produce graduates with better language skills to meet Japan's growing interaction with other countries. I feel that the process of selecting the criterion measures and constructing the test itself will produce a healthy dialogue about present programs and future goals among the faculty and result

in better teaching programs. Admittedly, it will cause extra work but I hope that I have shown some of the reasons why the benefits will far outweigh the efforts and make them worthwhile.

Bibliography

- Asano, Hiroshi and Johnson, Vance E., "The Measurement of Listening Comprehension Progress: An Experiment Using Data from Short-term Intensive Training." *Gaikokugo Kyoiku Ronshu*. No. 3. (The University of Tsukuba: Foreign Language Center, 1982) pp. 13-23.
- Burt, Marina, Dulay H., and Finocchiaro M. Eds. *Viewpoints on English as a Second Language*. New York: Regents Publishing Co., 1977.
- Carroll, John B., "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students." *Testing the English Proficiency of Foreign Students*. Washington, D. C.: Center for Applied Linguistics, 1961.
- Cartier, Francis A., "Criterion-Referenced Testing of Language Skills," *TESOL Quarterly*, vol. II (1968), pp. 27-32.
- Clark, John L. D., *Foreign Language Testing: Theory & Practice*. Philadelphia: Center for Curriculum Development, 1972.
- Cooper, Robert L., "An Elaborated Language Testing Model." *Language Learning* (Special Issue Number 3) 57-65, 1968.
- Davies, Alan, *Language Testing Symposium: A Psycholinguistic Approach*. London: Oxford University Press. 1968.
- Delattre, Pierre, "Testing Students' Progress in the Language Laboratory," *International Journal of American Linguistics* (Oct. 1960), pp. 77-92.
- Diederich, Paul B., *Short-cut Statistics for Teacher-made Tests* (Princeton, E. J.: Educational Testing Service, 1964) (Evaluation and Advisory Series).
- Dulay, H., Burt, M. and Krashen, S., *Language Two*. New York: Oxford University Press, 1982.
- Green, Jerald R. Ed., *Foreign-Language Education Research: A Book of Readings*. Vol. 11. Chicago: Rand McNally & Co., 1973.
- Harris David P., *Testing English as a Second Language*. New York: McGraw Hill, 1969.
- Hatch, Evelyn and Farhady, H., *Research Design and Statistics for Applied Linguistics*.

Rowley : Newbury House, 1982.

Heaton, J. B., *Writing English Language Tests : A Practical Guide for Teachers of English as a Second or Foreign Language*. London : Longman, 1975.

Jones, Randall L. and Spolsky, Bernard, eds., *Testing Language Proficiency*. Arlington, Va. : Center of Applied Linguistics, 1975.

Lado, Robert, *Language : The Construction and Use of Foreign Language Tests*. New York : McGraw-Hill, 1964.

Mackey, William F., *Language Teaching Analysis*. Norfolk : Longman, 1965.

—, "The Typology, Classification and Analysis of Language Tests." *Language Learning* (Special Issue No. 3) 163-166, 1968.

Madsen Harold S., *Techniques in Testing*. New York : Oxford University Press, 1983.

O'Brien, M. C. ed., *Testing in Second Language Teaching : New Dimensions*. Dublin : Association of Teachers of English as a Second or Other Language and Dublin University Press, 1973.

Palmer, Leslie, and Spolsky, Bernard, eds., *Papers on Language Testing : 1967-1974*. Washington : Teachers of English to Speakers of Other Languages, 1975.

Politzer, Robert L., *Linguistics and Applied Linguistics : Aims and Methods*. Philadelphia : Center for Curriculum Development, 1972.

Postovsky, Valerian A., "Effects of delay in oral practice at the beginning of second language learning". *Modern Language Journal*, 58 : 5-6, 1974.

—, "The priority of aural comprehension in the language acquisition process". *Proceedings of the Fourth International Congress of Applied Linguistics*. Stuttgart, 1975.

Rivers, Wilga M., *The Psychologist and the Foreign-Language Teacher*. Chicago : Univ. of Chicago Press, 1964.

—, "Listening Comprehension" in Donoghue, Mildred R. ed., *Foreign Languages and the Schools : A Book of Readings* Dubuque, Iowa : Wm. C. Brown Company, 1967, pp. 189-200.

—, "Listening Comprehension" *Teaching Foreign Language Skills*. Chicago : University of Chicago Press, 1968.

Savard, Jean Guy, 1968. "A Proposed System for Classifying Language Tests." *Language Learning* (Special Issue No. 3), 167-174.

Spolsky, Bernard, "Language Testing : The Problem of Validation" *TEOSL Quarterly*,

Vol. II (June 1968), pp. 88-94.

Spolsky, Bernard, *et al.*, "Preliminary Studies in the Development of Techniques for Testing Overall Second Language Proficiency," *Language Learning*, Special Issue No. 3(Aug. 1968), pp. 79-98.

Valette, Rebecca M., *Modern Language Testing: A Handbook* (New York : Harcourt, Brace, 1977.

Winitz, Harris and Reeds, James A., "Rapid acquisition of a foreign language (German) by the avoidance of speaking". *IRAL*. 11/4 : 295-317, 1973.

—, *Comprehension and problem solving as strategies for language training*. The Hague : Mouton & Co., Paris : N. V., Publishers, 1975.