

Survey Research on the Use of Placement Tests at Four-Year Universities in Japan

Yuko SHIMIZU

1. INTRODUCTION

In English education at the university level in Japan, a great deal of effort has been paid to improve its media, materials, curricula and even their administrative organization. As one example, some schools are organizing classes based on students' proficiency of English, presumably expecting efficient teaching and learning. In order to place the students into suitable levels of classes, English test scores on entrance examinations can be utilized. In many private universities in particular, however, all the newly enrolled students do not necessarily take the same examinations. Rather, their acceptance or rejection is determined based on more than one examination. It is not problematic if these different tests are parallel tests. As is often the case with many universities, however, the tests are not parallel in terms of their degree of difficulty and types of test items. Moreover, some students enter a university without having taken an examination. Therefore, it is not possible to obtain a common measuring scale to form sub-groups based on English proficiency.

Concerning English language teaching at the universities in Japan, Koike et al. (1983) conducted a large-scale survey on instructors of English in 1982. As a part of the results, they reported that 487 respondents out of 959 (50.8%) showed favorable attitudes toward dividing classes based on the level of the students' English proficiency. However, only 36 of them (3.8%) were actually engaged in such a teaching situation at that time (Koike, 1990).

Almost two decades have passed since the survey was conducted and various conditions both at

the social and educational levels have changed. Also, the number of research studies in language testing has notably increased. That means more than ever, teachers and researchers of English are becoming involved in measurement and evaluation in some way or other. In order to retrieve basic reference data of a testing situation in four-year universities in Japan, I conducted a questionnaire study in October 1999. The questionnaires were distributed to 616 four-year universities. Two hundred out of 616 questionnaires were answered and returned. That is, 32.5% of the respondents' data were analyzed. In this paper, I will first review the background of the study and then report on the findings obtained by the questionnaire. Finally feasibility and problems many universities encounter in developing and conducting English placement tests will be discussed in relation to their curriculum.

2. BACKGROUND

As far as we are concerned with classroom instruction in formal education, we have to be involved in evaluation whether we like it or not. The most immediate issue is evaluation of our learners through classroom activities, what is called classroom-based evaluation. In addition, more large-scale evaluation such as evaluation of a program, curriculum, school and organization are also a great concern for those who engage in administrative work. No matter what levels and scales we use, the evaluation we make must have a clear purpose.

Let us focus on language instruction, English in this case, to make the discussion more concrete.

Many language testers conceptualize types of language tests (Henning, 1987; Oller, 1979; Carroll, 1961). Although those attempts at conceptualizing and categorizing the tests are to facilitate our understanding of the multitude of purposes of different types of tests, ambiguity in terms of categorization may lead to misunderstanding. However, the distinction of criterion-referenced tests (CRTs) and norm-referenced tests (NRTs) enables us to understand roles of testing in a clearer way (Hudson & Lynch, 1984; Bachman, 1990; Brown, 1989, 1995). Two levels of evaluation as mentioned above are based on the types of decisions that need to be made—classroom-level and program-level decisions, which are explained by Brown (1996) in relation with CRTs and NRTs respectively. The former is to assess the amount of learning each learner has attained in a particular course, program or school. Achievement tests and diagnostic tests serve this purpose. The latter are to compare the performances of test takers and to give their relative positions. Examples of NRTs include entrance examinations, proficiency tests and placement tests.

Focusing on evaluation at the classroom level, the following figure, constructed based on Brown (1996, p9) and Carrol and Hall (1985, p9), will simplify our situation concerning testing in English programs.

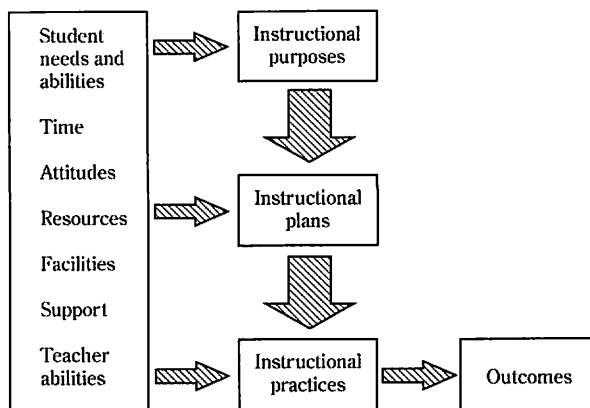
Figure 1 English Test in a Program

	Type of Decision	Type of Test	Type of Measurement	
Terminal Progress ↑↓ Placement	CRT (Classroom-level)	Diagnostic	Very Specific ↑↓ Very General	End-Of-Course
		Achievement		On-Course
	NRT (Program-level)	Placement		Pre-Course
		Proficiency		Enter

And furthermore, Genesee and Upshur (1996, p15) show the context of evaluation in the teaching-learning process. As Figure 2 indicates, classroom instruction consists of (1) objectives of instruction (purpose), (2) means for those objectives (plans) and

(3) activities taking place in the classroom (practices). They also refer to other factors such as students' needs and abilities, time, and resources as input factors, which have an effect on teaching and learning.

Figure 2 The Context of Classroom-based Evaluation



Talking about the relationships of teaching-learning and testing, it is essential to have the combined effects of these three factors: the testing system, the program itself and needs. However, the existing condition of each university is distant from such an ideal situation in more ways than one. Then, how effectively or ineffectively is testing functioning in reality? To find an answer to this question, I focused on the entrance period into an English program to see how testing is incorporated within the program.

3. THE STUDY

3.1 Purpose and Method

This study aimed at exploring the present condition and the problems of implementation of an English test, a placement test in particular. For the purpose of data collection, a questionnaire was designed to make an inquiry into the use of English language tests at four-year universities in Japan. Respondents were firstly asked if they were giving any large-scale English tests for general English

courses. Then the questions were targeted on placement tests, which is the main concern of the present study. (Note that *a large-scale test* in this paper refers to any test given to a whole body of students in a given situation, since no equivalent expression for *tōitsu/kyōtsū tesuto* is found in English.)

Those who answered in the affirmative to the question of using such large-scale English tests were then asked about the following points: the target population of the test (4.1.1), the time of the year to carry out the test (4.1.2), type(s) of tests being used (4.1.3), ways of utilizing the results (4.1.4), feedback to the students (4.1.5), and the administration of a placement test (4.1.6~4.1.8). In the result section of this paper, findings are presented in the above mentioned order.

Those who answered “no” to the question above were asked reasons why they do not give large-scale tests (4.2.1), past experiences and/or future plans (4.2.1) and the respondents’ perceptions of the ideal way to give such a test (4.2.2). Due to limitations of space, the results of those questions above will be briefly summarized in section 4.2 of this paper.

Finally, all the respondents were asked to make comments on English tests at the university level in Japan in relation to curriculum, instructional goals, test development and the use of standardized tests (4.3).

3.2 Samples

The questionnaires were distributed to 616 four-year universities: 95 national, 66 public and 455 private universities in October 1999. One thing I would like to add about the samples of this study is that the questionnaires were answered by individual teachers; some of their English curricula were at the school-wide level and others were departmental or divisional ones. Therefore, answers and opinions of the respondent do not necessarily represent those of their institutions.

4. RESULTS

Two hundred out of 616 questionnaires were answered and returned. That is, 32.5% of the responses were analyzed.

At the beginning of the questionnaire, all the respondents were asked if there were any large-scale English tests (excluding mid-term and final exams and quizzes given in a classroom level) which their school/department/division was administering in general English courses. Out of 200 respondents, 96 (48%) indicated they were conducting such English tests. Detailed information is shown in Table 1.

Table 1 “Does your institution use any large-scale English tests?”

Institution of higher learning	National university		Public university		Private university		Total	
	Yes	No	Yes	No	Yes	No	Yes	No
Responses	16	30	5	10	75	64	96	104
	34.8%	65.2%	33.3%	66.6%	54.0%	46.0%	48.0%	52.0%
Total (a) (ratio:b/a)	46 (48.4%)		15 (22.7%)		139 (30.5%)		200 (32.5%)	
Number of questionnaires sent (b)	95		66		455		616	

As mentioned above, the respondents’ answers varied depending on how English curricula were planned; some curricula would be inter-departmental or divisional and others not. Table 2 summarizes the number of respondents classified by department and the number and ratio of those who were administering English tests.

Table 2 Respondents Classified by Department

Departments	Number of respondents	Ratio in the whole	Yes	Ratio within the category
English*	34	17.0 (%)	26	76.5 (%)
Humanities**	31	15.5	16	51.6
Social Science	55	27.5	25	45.5
Education	12	6.0	1	8.3
Natural Science & Technology	19	9.5	9	47.4
Medical & Health	23	11.5	7	30.4
P.E. & Arts	8	4.0	4	50.0
Entire School	18	9.0	8	44.4
Total	200	100	96	48.0

*English (foreign language study and literature related to English)

**Humanities (humanities excluding English related fields)

Twenty-six cases out of 34 (76.5%), which were related to the fields of English, were administering

the English tests. The categories of humanities and social science, whose ratios were relatively high in the samples, also showed high ratios of administration of the tests. (51.6% and 45.5% respectively)

4. 1 Results of Those Who Give Some Large-Scale Tests

4. 1. 1 Who is Tested?

A. Scale of Test Administration

The target population of test administration was asked in the questionnaire. The respondents were to choose either (a) all the students in the entire school, (b) all the students in the department, (c) all the students in the division or (d) other. Table 3 summarizes the results.

Table 3 The Target Population of Test Administration

	(a)	(b)	(c)	(d)	total
English	7	2	12	5	26
Humanities	5	3	6	2	16
Social Science	10	10	0	4	24
Education	0	1	0	0	1
Natural Science & Technology	5	2	2	0	9
Medical & Health	6	0	0	0	6
P.E. & Arts	3	0	0	1	4
Entire School	4	3	0	0	7
Total (%)	40 (43.0%)	21 (22.6%)	20 (21.5%)	12 (12.9%)	93 (100%)

No response = 3

Since some respondents were from colleges that had only one department, departmental administration directly meant a school-wide level administration. Therefore, detailed and accurate information concerning the administration unit of the test could not be obtained in the present study. One tendency observed was, however, that tests were conducted in larger units than simply at the classroom or division level.

B. Students' Class Level

Table 4 shows the target class level of the test takers. No less than 92.1% of the respondents answered that they gave the tests to first year students. The higher the class level, the less these

English tests were given. This most probably is due to the general fact that most English classes are offered to underclassmen, i.e., first and second year students. Some schools gave the tests to more than one class level, which included the first year students.

Table 4 Class Levels to Administer the Tests

Class level	1st	2nd	3rd	4th
Single class level	59	3	2	0
More than one class level	12	6	5	
TOTAL	82 (92.1%)	27 (30.3%)	13 (14.6%)	5 (5.6%)

4. 1. 2 When is the Test Administered?

As to the frequency of test administration, 63.2% of the respondents answered that they administered them once a year, 32.2 % of them twice, and the rest three times a year. Among the 55 respondents, who administered once a year, 40 of them (72.7%) gave the tests at the beginning of the school year, that is in April. Also, irrespective of the number of times, 66 respondents (75.9%) reported administering the tests in April.

Table 5 Time to Administer the Tests (n=87)

Times per year	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Total
Once	40	2	4	0	0	1	0	2	2	2	2	0	55 (63.2%)
Twice	10	1	1	1	1	1	1	1	1	1	1	1	28 (32.2%)
	1	1	1	1	1	1	1	1	1	1	1	1	
	1	1	1	1	1	1	1	1	1	1	1	1	
	1	1	1	1	1	1	1	1	1	1	1	1	
	1	1	1	1	1	1	1	1	1	1	1	1	
	1	1	1	1	1	1	1	1	1	1	1	1	
	1	1	1	1	1	1	1	1	1	1	1	1	
Three times	1	1	1	1	1	1	1	1	1	1	1	1	4 (4.6%)
	1	1	1	1	1	1	1	1	1	1	1	1	
	1	1	1	1	1	1	1	1	1	1	1	1	
total	66	2	5	5	1	4	3	3	4	4	9	0	

4. 1. 3 What Types of Test are Administered?

Respondents were asked what types of tests were given to the students for such large-scale tests. Table 6 summarizes the results.

Table 6 Tests in Use (check all applicable items)

Originally-developed tests	52 (57.8%)		
Standardized tests	48 (42.2%)	TOEFL	21 (23.3%)
		TOEIC	13 (14.4%)
		Eiken (STEP)	8 (8.9%)
		CELT	2 (2.2%)
		G-TELP	1 (1.1%)
		Other	5 (5.6%)

Out of 90 respondents, more than half answered that they developed their own tests (52 respondents, 57.8%). (The sub-tests in those originally-developed tests are discussed below.) Some of those who had adopted using other standardized tests reported using various tests available in Japan. As shown in Table 6, the TOEFL and the TOEIC were relatively popular choices among others, 21 samples (23.3%) and 13 samples (14.4%) respectively. As to the STEP test (the Society for Testing English Proficiency), widely known as Eiken, grade levels used varied from the third grade to pre-first grade tests. Other tests included listening tests developed by the JACET (Japan Association of College English Teachers), the Michigan Test, and placement tests related to textbooks used in class.

4. 1. 3. 1 Originally-developed tests

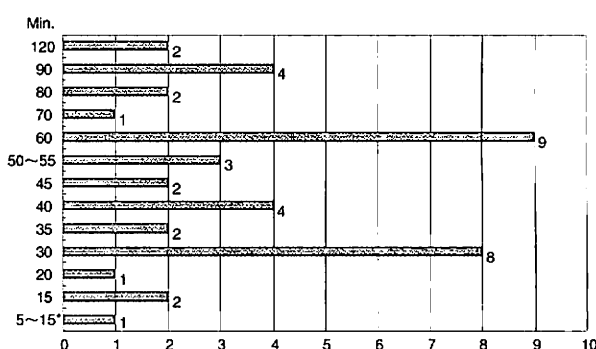
What kinds of tests have been developed for the use of their own program? Forty-five respondents provided information regarding time allotments and test formats. I will briefly summarize the information obtained.

(a) Time allotment

As seen in Figure 3, one school said they spent only 5 to 15 minutes for the test. (see * in the figure) However, this school reported giving oral interviews, using that amount of time for each student. Others had reported administering tests to a group of people at the same time, spending anywhere from 15 to 120 minutes. However, time allotments of 30 to 60 were

the most popular. This is probably because they want to complete the tests within one class period, which is usually 90 minutes at most universities in Japan.

Figure 3 Time Allotment

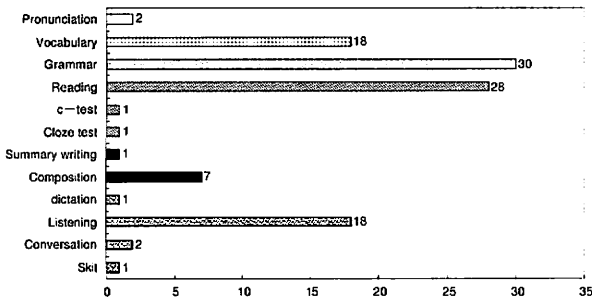


(b) Sub-tests in use

One set of tests often contains several sub-tests. For example, the TOEFL-ITP has three sub-tests: listening comprehension, structure & written expressions and reading comprehension. However, two respondents answered that they only gave a grammar test, three respondents gave a listening test and one gave an oral interview. The rest of them (39 samples) were giving two to five sub-tests in their test sets. Figure 4 summarizes sub-tests used in their test. Grammar (76.9%) and reading comprehension (71.8%) were most popular followed by vocabulary (46.2%) and listening comprehension (46.2%). These sub-tests are common and widely used on standardized tests as well.

As to test formats, some reported using summarization and compositions, which require extended production responses of the test takers. Most of the tests were, however, recognition types such as a multiple-choice test and recall types such as a completion test and a re-arrangement test, all of which were able to be evaluated objectively. These test types make it possible to carry out machine scoring and many of the respondents reported to already having adopted card readers.

Figure 4 Sub-tests in Use



4. 1. 4 How are the Test Results Used?

The results of tests can be applied in various ways. How the test results were actually used is the topic of this section. An actual question in the questionnaire is as follows:

Generally speaking, test results can be used as pre- and post-test tools to examine teaching effects, etc. In your situation, how do you use the results of the test you indicated in questions above?

Hashimoto (1977) indicates that evaluation in a teaching-learning situation in school has four major functions: teaching, learning, administrative and research functions. The responses in this study involved all those functions; however, many aimed at facilitating teaching and learning in particular. Table 7 shows that as many as 57 respondents (64.0%) used the tests for a placement purpose, including three of whom used the tests only for speaking classes. Since many administered the tests at the beginning of their school year (see Table 5), it is understandable that they used the tests for the purpose of forming relatively homogeneous classes in terms of learners' abilities and language competencies. Other than a placement purpose, the results were used to grasp the learners' English ability (13.5%) and to examine effectiveness of teaching (14.6%) to utilize them as data and reference for teachers.

Table 7 Ways to Use the Test Results (n =89)

For a placement purpose	57 (64.0%)
Grasping English ability	12 (13.5%)
Measuring achievement	2 (2.2%)
As part of school record	5 (5.6%)
To test teaching effects	13 (14.6%)
As part of requirement	6 (6.7%)
Data for enrollment of graduate school, etc.	3 (3.4%)

4. 1. 5 Feedback to the Students

Concerning feedback to the test takers, 46.2% (36 out of 78 respondents) reported providing it and 42 respondents (53.8%) answered that they did not give any form of feedback to the test takers. Twenty out of those 42 who did not provide feedback were using the results for placement purposes. Test results sometimes need to be kept confidential for administrative purposes. However, the results can possibly become a motivational tool for the learners and provide diagnostic information to both learners and teachers. Whether or not we give the feedback and how we give it will be crucial issues to discuss in administering such tests.

4. 1. 6 Administration of Placement Tests

So far all the questions were concerning a large-scale test. In this section of the questionnaire, however, respondents were directly asked about English tests for placement purposes. Results showed that 64 respondents out of 96 (66.7%) conducted the test as a placement test, while twenty-six used the test for other purposes (see also Table 1). That is, approximately 70% of those who were giving some English tests to their target student groups used the results to identify the levels of their English ability to place them in appropriate classes or courses. (No responses=6, 6.3%)

In the next sections of 4.1.7 and 4.1.8, I will present the results of those who did not conduct the test for a placement purpose and then those who did.

4. 1. 7 Those Who Do Not Administer a Test for a Placement Purpose

Twenty six respondents who conducted some large-scale English tests but did not use them as placement tests were asked to answer the following open-ended questions: A. Reasons why they do not give a placement test and B. Past experiences in and future plans for conducting placement tests.

A. Reasons for Not Administering a Placement Test

Some respondents admitted that they needed to form their English classes by the learners' English abilities if they took the immense number of students, their abilities and the curriculum into consideration. However, there were some practical conditions that prevented them from conducting the tests. Those conditions were partially from the lack of consensus among teachers on such a class formation and the burdens of clerical and administrative tasks.

B. Past Experiences and Future Plans for Conducting Placement Tests

Five out of 15 respondents (33.3%) who did not use placement tests at present had experienced conducting them in the past. Eight out of 18 (44.4%) respondents were planning to start giving the tests, three of whom had already started to prepare for them.

4. 1. 8 Those Who Give Placement Tests

Those who conducted some large-scale English tests and used them for placement purposes were asked to write their opinion about the following aspects: A. How they form classes, B. Advantages and disadvantages of the tests now in use and C. Opinions of the tests used for this purpose.

A. How to Form Classes

Among the forty-five respondents who wrote some comments on this item, one did not use the test

result to form any levels of English classes. Another used it as secondary resource and gave students' own opinions first priority. The rest of the respondents said they divided the classes into two to ten levels based on the test results. Table 8 summarizes the results.

Table 8 Class Levels

Number of levels	Frequency	Note
2	15	<ul style="list-style-type: none"> • To select the upper group. (n=4) • To choose first hundred students to form upper level classes irrespective of their departments, then the rest of the classes are formed. (n=5) • Grading is used only for listening classes. • Grading is used to select the upper group for the second year reading classes.
3	20	• Grading is used for conversation classes.
4	3	—
5~6	3	—
8	1	—
10	1	—

B. Advantages and Disadvantages of the Tests in Use Now

The responses to this item were roughly divided into three categories: (a) teaching, (b) test administration and (c) the tests themselves.

(a) Teaching

Placement tests are to place the students in the appropriate levels of class, which will supposedly allow teachers to provide more appropriate and effective instruction. Making full use of students' levels of English ability, some institutes were using shared textbooks or developing their own textbooks to match the learners' needs and English abilities, which was one of the main advantages pointed out. Others pointed out that it was an advantage for both students and teachers to be given feedback of the tests as diagnostic information, which would hopefully promote a favorable learning-teaching climate.

One of the prominent disadvantages pointed out by many respondents was the psychological impact to lower level students. Those respondents expressed concern that placing the students in the lower level

classes might possibly give a sense of inferiority and diminish their motivation to learn. Some respondents had doubts whether or not we could provide appropriate teaching to meet the assigned students' levels. One respondent mentioned that some teachers might make the mistake of misinterpreting the test results as the learners' overall abilities and their motivations.

(b) Test Administration

The ease of scoring the tests becomes a crucial and realistic issue when we conduct any test on a large scale. Moreover, results have to be obtained promptly to form classes and start lessons in case of the placement test. Most of the respondents who reported that it was an advantage of the tests to measure a large number of students at the same time were using card readers to mechanically score the tests. However, some claimed that it was a burden to spend time conducting and scoring the tests during the beginning of the school year, which was the busiest time of the year, to say nothing of the burden of developing their own tests. In the case of using standardized tests on the market, the time it took to receive test results was an important factor to take into consideration.

Some referred to the problem of students who did not take the test on the assigned day for some reasons. The teachers have to use some other source of information to place the absentees, which may often cause some administrative problems.

(c) Tests Themselves

The major advantages of giving the same tests at the same time indicated by many respondents were to be able to objectively measure learners' English ability, to measure the students on the same scale and to obtain all the students' data. Also some said they could utilize the tests to grasp the students' progress during the course and the change in the traits of entering students longitudinally. Some of the

universities that were developing their own original tests were constantly revising the tests to refine them for their particular student body. Those who were using commercially recognized tests took it as an advantage that the scores were also recognized by a larger community.

There were, however, several negative aspects pointed out. Firstly, in the case of an originally-developed test, objectivity, validity and reliability of the test came into question. Analysis would be needed to see if the test in use was the most appropriate tool to measure their students' ability. In reality, there may be few institutes that are in favorable environments to develop and analyze the tests. Secondly, when they use other standardized tests, many reported facing some problems in terms of precision—particularly for the lower level students. Also feasibility in terms of difficulty and the skills that such tests are trying to measure and content validity of the tests with actual contents and skills the students encounter in their English courses were issues that were raised.

C. Tests Used for Placement Purposes

Those who were giving placement tests were asked about the types of tests in use. Fifty-five (90.2%) were using the tests they answered in 4.1.3 as a placement test. However, six of them (9.8%) were using different measuring tools from the ones used as large-scale tests. Among those six respondents, one was using the TOEFL but the other five were developing their own placement test. That is, as

Table 9 Tests Used for Placement Purposes

		Using the same test as 4.1.3	Using a different test from an answer in 4.1.3	Total
Originally-developed tests		31	5	36 (59.0%)
Standardized tests	TOEFL	10	1	11 (18.0%)
	TOEIC	6	0	6 (9.8%)
	STEP	4	0	4 (6.6%)
	CELT	0	0	0 (0.0%)
	G-TELP	0	0	1 (1.6%)
	others	3	0	3 (4.9%)
Total		55 (90.2%)	6 (9.8%)	61 (100%)

shown in Table 9, nearly 60% were developing their own tests to use for a placement purpose. It can be concluded that many schools are developing their own tests to meet the conditions and demands of their schools—students' levels of English, purposes of learning, cost and time.

4.2 Results of Those Who Do Not Give Any Large-Scale Tests

4.2.1 Reasons, Past Experiences and Future Plan

One hundred and four respondents out of 200 answered that they did not conduct any English tests on a large scale. Ninety-eight of them gave reasons why they did not give such tests: 34.7% (34) said that it was unnecessary and 15.3% (15) that it was impossible to conduct. However, it is worth paying attention to the fact that 17 schools (17.2%) had conducted placement tests in the past, seven of which were planning and/or hoping to conduct placement tests in the future. As Table 10 shows, approximately one-third of the respondents were in the planning stage of conducting test administration.

Table 10 Past Experience and Future Plan of a Placement Test

	Past experience	Future plan
Yes	17 (17.2%)	31 (31.3%)
No	82 (82.8%)	68 (68.7%)
Total	99 (100%)	99 (100%)

4.2.2 Opinions on What Constitutes an Ideal Test

Assuming that they were to give some tests on a large scale, they were asked questions concerning scale, students' class levels, the amount of time given to conduct the tests, etc. Detailed results will be summarized in a research report submitted to the Ministry of Education (Shimizu, 2001). Here I summarize the types of tests to use, purpose of the tests and opinions about placement tests.

A. Types of Tests to Use

Table 11 summarizes the responses concerning the tests that respondents would like to use. Approximately 70% of the respondents showed intention of using standardized tests, contrary to the fact that those who were already giving such tests tended to use originally-developed tests (see Tables 6 and 9). The figures in Table 11 show that the test of choice concentrated in the highly recognized tests in Japan such as the TOEFL, the TOEIC and the STEP, 34.6%, 34.6% and 29.5% respectively. As for other tests, the listening comprehension tests developed by the JACET, the SLEP (Secondary Level English Proficiency Test), and Kohgyo Eiken were also named.

Table 11 Tests to Use n=78
(Check all applicable items)

Originally-developed tests	24 (30.8%)		
Standardized tests	54 (69.2%)	TOEFL	27 (34.6%)
		TOEIC	27 (34.6%)
		STEP	23 (29.5%)
		CELT	1 (1.3%)
		G-TELP	2 (2.7%)
		Others	9 (11.5%)

B. How to Use the Results of the Tests

Seventy respondents reported on how the results of tests should be used. As seen in Table 12, as many as 30 respondents out of 72 were interested in using the results for the purpose of placement.

Table 12 Ways to Use the Test Results
(Check all applicable items)

As a placement test	30
Grasping English ability	12
Measuring achievement	2
As a part of school record	5
To test teaching effects	15
As part of requirement	10
Referential data for teaching	17

Although they used different ways to express this, it was clear that many were interested in using the results to facilitate teaching.

C. Opinions on Placement Tests

Many respondents showed an interest in placement tests and many opinions in both negative and positive aspects. Views from the following points were summarized below: (a) problems in administration, (b) negative factors towards learners, (c) appropriateness of the tests, (d) positive opinions and (e) ways to organize classes.

(a) Problems in Administration

It is said that there has recently been a decline in the levels of academic attainment of university students. Under such conditions, many respondents recognized that a placement test is necessary to define teaching objectives and to evaluate the effectiveness of teaching. However, many disadvantages were pointed out, e.g., the difficulty to obtain a certain time allotment to administer the tests, the burdens coming from test development, test administration, data processing and management and the division of responsibility. Moreover, some pointed out the problems of the current educational climate in Japan and of forming classes based on students' abilities.

(b) Negative Factors towards Learners

Efficiency of class management was recognized by many respondents as one advantage of class organization by levels. At the same time, however, it was recognized as possibly having a negative impact on the students. Many respondents reported that they could easily see higher-level students and highly motivated ones to benefit from such teaching. They pointed out, however, that it was risky for lower-level students in that it might foster negative attitudes towards the English language and English learning. Also, one pointed out that homogeneous classes in

terms of the students' levels of English would decrease the number the opportunities for students to learn together and help learners of different proficiency levels to learn from each other.

(c) Appropriateness of the Tests

Many respondents mentioned appropriateness of the measuring instruments they used. For example, some said that many standardized tests were often out of touch with reality in terms of their difficulty and content. Furthermore, these tests might not necessarily be reliable for some target population and also could not measure students' motivation to study. Some complained that these tests would not be able to measure proficiency levels either. Idealistically speaking, it is necessary to develop criterion-referenced tests to measure what the learners learned throughout the program.

One respondent stated that unless the validity of the originally-developed tests were socially acknowledged, evaluation of English education at the university level would remain unrelated to the larger societal needs.

(d) Positive Opinions

Some respondents expressed the urgent need for the use of placement tests. Such universities whose students were diverse in terms of their English abilities felt a strong need for efficient teaching and learning and for the appropriate selection courses/classes to take. Others regarded test results as important information sources just like clinical records. Some viewed the test as a positive light for lower-level students. That is, the students could learn at their own pace.

(e) Ways to Organize Classes

Some ways for organizing classes mentioned in the responses were to form classes based on students' proficiency of English not for the whole English curriculum but for certain subjects and to

regroup the classes regularly after each semester, for example. Furthermore, one implied to give the students priority to decide their classes to take.

4.3 English Tests and Curriculum

At the end of the questionnaire, respondents were asked to write freely about curriculum, test development and English education in university and an English test as it ought to be. Eighty-seven respondents stated their opinions about various aspects. Analyzing their opinions, the responses were divided into the following categories: objectives, test development, utilization of standardized tests and the importance of testing.

4.3.1 Objectives to Attain

It is worthy of special mention that twenty-five out of 87 respondents (28.7%) referred to learning and teaching objectives, all of whom expressed dissatisfaction with the lack of clear objectives and need for making them clear. Some said, by clarifying goals and English abilities that were needed, they could make appropriate decisions in terms of what type of curriculum to follow, the appropriate textbooks to use and the type of tests to measure their learners. Although it was a minority opinion, some pointed out that conducting a large-scale test at school would make it possible to change teachers' attitudes toward curriculum and evaluation. Regarding setting up teaching objectives, more than one respondent worried about the difficulty of achieving consensus among teachers, which mainly came from different views on students' 'English ability.'

Various students coexist as learners in each university. Their diversity comes from not only their specialties and interests concerning English but from their abilities in English. The latter aspects are being bipolarized in many schools, even in the same faculty or the same curriculum. Therefore, more than one separate objective based on the students' level of

English may be needed. Schools which focus on EAP(English for Academic Purposes) and ESP (English for Specific Purposes) will surely conduct needs analysis in a many-faceted way.

4.3.2 Test Development

Those who had already been involved in developing original tests touched on realistic problems such as the difficulty of constructing the tests and the physical burden in administrating the tests. Some of them reported a need for developing an ideal test for given situations as well as standardized tests which are appropriate for Japanese learners of English. These needs were also strong among those who did not utilize some type of large-scale English test nor use a placement test. Many were expecting a standardized test that could be used on a nationwide scale in particular. That implies that within the present state of English education in Japan, there is no satisfactory standardized test for learners at the university level. Concerning such a test, one respondent suggested that several universities cooperate to develop a test that would become socially accepted.

Moreover, some focused on oral proficiency tests. This is due to a movement that attempts to avoid abuses of paper-pencil type objective tests, i.e., cheating and measuring the test taker's abilities in the target language use situation.

4.3.3 Utilization of Standardized Tests

As indicated above, there is a strong need for the creation of an appropriate standardized test. Among the responses, some pointed out advantages and disadvantages of commercially available tests at present. The majority referred to the TOEIC, the TOEFL and STEP and they said that encouraging students to take those tests would strengthen their motivation and become a stimulant to study the language of their own accord. Possible ways to use the scores of those tests mentioned by some

respondents were to grasp the learners' ability and their gains and to use them as data to design and redesign a curriculum and program. Compared to some originally-developed tests, those standardized tests are reliable, which is another advantage. In addition, it becomes an advantage for the students to grasp their English ability based on a widely-accepted standard.

On the contrary to those positive opinions, some had misgivings about using standardized tests. For example, some pointed out that to increase scores might become the main and only purpose of instruction. Another mentioned that teachers and/or schools did not seem to be acting on their own initiative and they totally depended on the concern of the organizers of the commercially produced tests.

Since standardized tests do not have a direct relationship with curriculum in many cases, those tests cannot be used as an achievement test, which is clearly a disadvantage. Thus, it is ideal to pursue two lines of testing: some appropriate and inexpensive standardized tests, i.e., NRT as well as teacher-made tests based on the curriculum and objectives in a given situation, i.e., CRT.

How the TOEIC and the TOEFL fit into curriculum is now under review at several universities. Their interest is particularly in whether or not they give credit to students who obtain a certain number of points. Obviously there are a lot of issues to discuss concerning those tests. Among the responses, I found that many teachers are hoping in the near future that some standardized English tests will be developed by some official organization, in order to use their scores as a requirement for graduation, conferring credits, data for career-related issues and for research purposes in the field of testing.

In addition to the standardized general English tests, there is a need for English tests for specific purposes such as pharmaceutical English, since English education is seen as a mediator for

specialized subjects these days.

4. 3. 4 Importance of Testing

One respondent made a suggestion that we should reconsider testing not based on teachers' intuitions or experiences only but on theory and practice. It is regrettable that not all universities in Japan have language testers on staff as teachers of English. Therefore, more than one specialist in language testing in several universities need to work together to develop tests and discuss how measurement and evaluation should be undertaken. With many people being involved in the field of testing, hopefully the right decisions on the use of measurement tools will be made and test development will be carried out with confidence in the near future.

5. CONCLUSION

Analysis of the data in the current study indicates that many universities were interested in and planning to give a placement test or already administering such a test. Among the ninety-six who were administering large-scale tests at present, 64 of them were using the test results for placement purposes and eight of them were planning to use the test for placement purposes. Although 104 out of 200 answered that they were not giving any large-scale tests to their students, 31 of them are considering giving the tests, which means that the population of people involved in placement tests is potentially large and many are taking an increasing interest in it.

Some universities reported developing their own placement tests, which suggests a real need and stresses the importance of original measurement tools to serve their own curriculum and English levels of the target learners. There were realistic problems to be confronted, however. One of them was a practical issue including the lack of human resources and time to develop and administer such

tests. Another problem was that there is no established procedure or theoretical foundation for test development nor a guideline for statistical analysis that each school should follow.

Using various types of tests in the process of teaching and learning, we make many decisions that have an impact on our learners. In this sense, we take on the enormous responsibility for the students. Therefore, we have to deliberate carefully on measurement and evaluation, giving thought to influences of tests on individuals—both teachers and students and society.

When we are involved in testing, we must first make the purposes of the test clear before we decide on the appropriate measuring tool to use. CRESST (the National Center for Research on Evaluation, Standards, and Student Testing), which is a research body of measurement and evaluation for elementary and middle schools in the USA, enumerates the essential elements a test has to possess. They are as follows: fairness, content quality, content coverage, meaningfulness, cost, and efficiency. Those elements are congruent with many of the aspects pointed out in the present study. In order to find a more suitable testing tool or tools under the educational conditions present in each individual university, it is a must to work as a team of people, schools and organizations.

The present survey was conducted on four-year universities concerning English tests and showed only one aspect of English education in Japan. However, many common problems were observed across the universities. Those problems becoming points of contact, some opportunity to learn from each other will hopefully be formed in the near future.

Notes: The present study was based on a 2-year research project financially supported by the Ministry of Education Grant-in-Aid for Scientific Research (c)-(2). Also, this research paper is a revision of a presentation given by the author at the

JACET Convention in Okinawa, in November 2000.

REFERENCES

- Aoki, S. (ed.) (1985). *Eigo no hyoukaron*. Taishuukan-shoten.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University press.
- Bachman, L. F. and Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.
- Brown, J. D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly*, 23, 65-83.
- Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. New York: Heinle & Heinle Publishers.
- Brown, J. D. (1996). *Testing in language programs*. Prentice Hall Regents.
- Carroll, J. B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. In *Testing the English proficiency of foreign students*. Washington, DC: Center for Applied Linguistics:30-40.
- Carroll, B. J. and Hall, P. J. (1985). *Make your own language tests*. Pergamon Press.
- CRESST (National Center for Research on Evaluation, Standards, and Student Testing) . from the Word Wide web: <http://www.cse.ucla.edu/>
- Genesee, F. & J. A. Upshur. (1996). *Classroom-based evaluation in second language education*. Cambridge University Press.
- Hashimoto, S. and Hidano, N. (1977). *Kyouiku hyouka no kangaekata*. Toshokan.
- Henning, G. (1987). *A guide to language testing*. Newbury House.
- Hudson, T. & Lynch, B. (1984). Criterion-referenced approach to ESL achievement testing. *Language Testing*, 1, 171-201.
- Ikeda, H. (1976). *Chousa to sokutei*. Shinyou-sha.
- Ikeda, H. (1992). *Tesuto no kagaku*. Nihon Bunka Kagaku-sha
- Koike, I. et al. (1983). General survey of English language teaching at colleges and *universities in Japan (I)*. Daigaku Ippan Eigo Kyouiku Jittai Chousa Kenkyukai.
- Koike, I. et al. (1990). *A General survey of English language in Japan*. Eigo Kyouiku Jittai Chousa Kenkyukai.
- Oller, J. W. Jr. (1979). *Language tests at school*. Longman.
- Shimizu, Y. (2001). *Survey research on the use of placement tests at four-year universities in Japan and a preliminary study for item banking*. Report of Research Project, Grant-in-aid for Scientific Research (c) (2), project number 11680286.
- Wada, M. (1999). *Gengo tesuto no kiso chishiki*. Taishuukan-shoten.