

# **The problem of Japanese High School Midterm and End of term Tests**

**Kenichi Otsuka  
(Waseda University)**

## **1. Introduction**

Most senior high school students in Japan need to pass five term tests to get credits and teachers who are in charge of the classes make these tests. To take a good score is significant for all students because grades depend on test scores. However we see tests that are badly organized and do not measure language abilities in a proper way. It is obvious that such tests cause a negative backwash effect to learners. The purpose of this research is to investigate the problems of midterm and end of term English tests that are used at Japanese public senior high schools.

## **2. Method**

### **2.1 Tests Data**

Several English subjects are taught at Japanese senior high schools and most schools select “English I” and “English” II as compulsory subjects rather than “Oral Communication I and II”. “English I” is taught for freshmen and usually juniors and seniors study “English II”. The main objective of these two is to develop students’ listening, speaking abilities as well as reading and writing ones. For this research, fifteen sets of “English I” and “English II” term tests were collected from six different general course public senior high schools in a certain prefecture. The rank of the schools varies from the top to the lowest. All these tests were actually used this year and created by teachers who were in charge of each class. These tests were conducted within 50 minutes in a classroom and the full mark was 100. Six schools were ranked from A to F by the researcher according to an unofficial data of public high school entrance examination made by preparatory schools for high school entrance examinations. A is the top school whose students could be regarded as a high level learners of all subjects and F could be said to be the opposite to this.

### **2.2 Investigation of what is tested and what items are used most**

To investigate what was tested, eight categories were selected in terms of the testing point; Vocabulary, Translation, Grammar, Reading Comprehension, Pronunciation, Accent, Listening Comprehension, and Other. In this research Writing was included in Translation, because there were no items which could elicit students’ writing productive skills such as “...Write your opinion about this issue in English.” And all items which asked students write English sentences were made with Japanese translated sentences. These items were regarded as Translation from Japanese to English. (See Appendix 1 No.19) And then twenty-seven types of item were selected to study what items were used most. The examples of each category are shown in appendix 1.

### **2.3 Pointing out inappropriate items and styles**

All items were analyzed and inappropriate test items or question styles were pointed out in terms of test writing techniques introduced by Alderson et al. (1995),

Hughes (2003), Shizuka (2002), and Wakabayashi, and Negishi (1993).

## **2.4 Item analysis**

One test was selected and the items were analyzed. To see if each of the items were good or not and the test are reliable or not, reliability, item difficulty, and item discrimination power were examined.

## **2.5 Correlation between term test and another test type**

To study what does the score of the term test mean, a correlation between a term test and proficiency test was examined.

## **3. Results and discussion**

### **3.1 What was tested? And what item was used most?**

From Table 1.1, it is obvious that translation and grammar were mostly tested in every school. Due to the government policy, Japanese English teaching style is supposed to shift from grammar translation method to more communicative emphasized way but regrettably, there is only one school that tests listening comprehension. It is said that term tests reflect daily classroom activities. According to a researcher's interview with each test maker, the collected tests were based on their own text book and what teachers taught in their classes. Besides, no interview test was conducted. This result implies that English listening or speaking abilities are not being taught in "English I" and "English II" classes so much. It means translation and grammar are still considered to be important by many teachers. The evidence of this also can be seen in Table 1.2. It reveals clearly that items related to "translation" such as Partial Translation Gap Filling (PTGF) are preferred by many test makers.

Example of PTGF is following;

Q: Fill in the blanks to make the sentence match Japanese meaning. (Written in Japanese originally)

1 I sat ( ) at the beautiful sunset. (A Japanese translation comes here.) answer : "looking"

2 He remains ( ) (A Japanese translation comes here.) answer : "seated"

According to a teacher who wrote this item, this question was designed to measure not a skill of translation or vocabulary but a grammatical skill. We are not concerned here with the question of the theoretical background but from the teachers' point of view, PTGF is convenient to use because it is easy to mark and moreover, the idea behind it is that partial translation is beneficial to learn not only translation but also vocabulary and grammar.

The result shows that teaching at classroom has still placed emphasis on its grammar and translation rather than communication.

### **3.2 Inappropriate test items and styles**

Each item was investigated from the point of view of item writing techniques. As a result, some problems such as layout, distracters, instruction, and testing points could be found from the collected tests. Some examples are following.

<Problem of a layout>

There is a test which text and questions were printed separately on both side of a sheet of paper. Students need to turn the paper when they want to read text or questions. This kind of problem should be modified by using an extra sheet of paper.

<Problems of distracters>

Example;

Q: Choose 1, 2, 3, or 4 which is closest in meaning to a).

- a) I don't like traveling in a group very much.
- 1, (A sentence written in Japanese comes here.)
  - 2, (A sentence written in Japanese comes here.) (= I hate traveling in a group.)
  - 3, (A sentence written in Japanese comes here.) (Correct answer)
  - 4, (A sentence written in Japanese comes here.)

In this case, distracters 1 and 4 are extremely bad because these two sentences obviously do not make sense at all in Japanese. Distracters should be attractive to at least some students but this time, they weren't.

<Problems of instruction and matching item>

Example;

Q: Match a-f on the left with 1-8 on the right to make each sentence describe the story of Lesson 2.

- |  |   |
|--|---|
| a, Hoshino Michio                      | 1, is minus 40 degrees Celsius.                                 |
| b, The mayor of Shishmarf              | 2, lived in Alaska and took photos.                             |
| c, The winter in Alaska                | 3, are part of the great Mother Earth.                          |
| d, The temperature in winter in Alaska | 4, gives people silence and sometimes gives animals a new home. |
| e, All living things                   | 5, call the polar bears "Nahook".                               |
| f, The snow                            | 6, lasts for half of the year.                                  |
| g, Eskimos in Alaska                   | 7, means the kind of the ice.                                   |
| h, "Nanook"                            | 8, sent the answer to Mr. Hoshino.                              |

The question asked students to make each incomplete sentence complete and these complete sentences need to match the content of a story in meaning. However, the original text of the story was not printed on the paper. Consequently, students had to recall the whole story without looking it to answer this question. This would have been possible if students were to measure memory retention, however this item should be modified because the purpose of this item was to elicit students' reading comprehension.

In addition to this, there is another problem in this question. From a grammatical point of view, some incomplete sentences in the left side match several choices from the right. For example, a sentence [a, Hoshino Michio] could make sense when it matches with 1, 2, 3, 4, 5, and 8. This causes students confusion to make a decision. To avoid unstable measurement, this question should be rechecked and modified.

<Problem of presentation>

Example;

Q: Choose a number of the word which [ ] is pronounced differently.

(1) 1, en[ou]gh 2, b[o]ttle 3, [u]ncle 4, bec[o]me

(2) 1, arr[i]ve 2, dr[i]ft 3, exc[i]ted 4, inv[i]te

In this case, the brackets [ ] were used to distinguish syllables which have the primary stress, however these brackets [ ] usually stand for phonetic symbols. Since 1-4 in each question are not phonetic symbols, the test maker should avoid using.

		School name																	
		English I						English 2											
Testing point		A	B	C	D	E	F	A	B	B	C	C	D	E	E	F	Total		
1	Vocabulary	23.5%	9.2%	5.3%	0.0%	19.6%	37.7%	0.0%	15.3%	15.8%	9.4%	12.1%	13.2%	9.7%	12.3%	26.3%	14.1%	<b>1 Translation</b>	<b>31.0%</b>
2	Translation	52.9%	18.5%	19.3%	27.6%	32.6%	6.5%	47.9%	30.5%	31.6%	32.1%	34.8%	23.7%	32.3%	42.1%	38.6%	31.0%	<b>2 Grammar</b>	<b>27.3%</b>
3	Grammar	15.7%	30.8%	31.6%	34.5%	34.8%	49.4%	1.4%	27.1%	24.6%	11.3%	13.6%	36.8%	45.2%	29.8%	24.6%	27.3%	<b>3 Vocabulary</b>	<b>14.1%</b>
4	Reading Comprehension	7.8%	20.0%	19.3%	19.0%	2.2%	0.0%	2.8%	10.2%	10.5%	28.3%	22.7%	10.5%	12.9%	15.8%	8.8%	12.6%	<b>4 Reading Comprehension.</b>	<b>12.6%</b>
5	Pronunciation	0.0%	6.2%	10.5%	8.6%	8.7%	0.0%	0.0%	8.5%	8.8%	9.4%	1.5%	7.9%	0.0%	0.0%	0.0%	4.3%	<b>5 Accent</b>	<b>6.3%</b>
6	Accent	0.0%	15.4%	14.0%	10.3%	0.0%	0.0%	11.3%	8.5%	8.8%	9.4%	7.6%	7.9%	0.0%	0.0%	0.0%	6.3%	<b>6 Pronunciation</b>	<b>4.3%</b>
7	Other	0.0%	0.0%	0.0%	0.0%	2.2%	6.5%	25.4%	0.0%	0.0%	0.0%	7.6%	0.0%	0.0%	0.0%	1.8%	3.4%	<b>7 Other</b>	<b>3.4%</b>
8	Listening Comprehension	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	11.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	<b>8 Listening Comprehension.</b>	<b>0.9%</b>

**Table 1,1 <What was tested?>**

		English I						English 2											
Item name		A	B	C	D	E	F	A	B	B	C	C	D	E	E	F	Total		
1	A Accent	0.0%	14.9%	14.0%	10.3%	0.0%	0.0%	11.3%	8.5%	8.8%	9.4%	7.6%	7.9%	0.0%	0.0%	0.0%	6.3%	<b>1 PTGF</b>	<b>22.4%</b>
2	C Correction	3.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	<b>2 OT</b>	<b>8.5%</b>
3	CV Conversion	0.0%	0.0%	0.0%	0.0%	0.0%	11.7%	0.0%	8.0%	1.8%	9.4%	0.0%	7.9%	1.6%	0.0%	0.0%	2.2%	<b>3 BGF</b>	<b>8.3%</b>
4	BGF Banked-Gap filling	0.0%	14.9%	17.5%	20.7%	0.0%	26.0%	0.0%	8.8%	0.0%	1.5%	0.0%	8.1%	8.8%	0.0%	8.3%	8.3%	<b>4 T1</b>	<b>7.2%</b>
5	GF-COR Correction Gap filling	0.0%	9.0%	7.0%	0.0%	0.0%	0.0%	0.0%	6.8%	7.0%	0.0%	0.0%	7.9%	9.7%	7.0%	3.5%	3.8%	<b>5 SA</b>	<b>6.7%</b>
6	GF Gap filling	0.0%	11.9%	0.0%	0.0%	10.9%	0.0%	0.0%	1.7%	7.0%	0.0%	1.5%	0.0%	1.6%	0.0%	12.3%	3.1%	<b>6 A</b>	<b>6.3%</b>
7	IT Information transfer	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.6%	0.0%	0.0%	0.0%	0.1%	<b>7 MC</b>	<b>6.1%</b>
8	M Matching	0.0%	0.0%	0.0%	13.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	15.2%	21.1%	0.0%	0.0%	0.0%	3.0%	<b>8 T</b>	<b>5.0%</b>
9	MC Multiple choice	5.6%	0.0%	1.8%	0.0%	19.6%	0.0%	0.0%	1.7%	1.8%	22.6%	22.7%	0.0%	19.4%	0.0%	0.0%	6.1%	9 T/F	4.9%
10	OT Ordering task	5.6%	6.0%	5.3%	13.8%	6.5%	9.1%	1.4%	15.3%	12.3%	1.9%	1.5%	10.5%	12.9%	14.0%	14.0%	8.5%	10 OTHER	4.1%
11	P Paraphrase	1.9%	6.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.7%	1.8%	1.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	11 GF-COR	3.8%
12	PR Pronunciation	0.0%	0.0%	0.0%	0.0%	6.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	7.9%	0.0%	0.0%	0.0%	0.7%	12 T2	3.5%
13	PGF Paraphrase Gap filling	0.0%	0.0%	0.0%	0.0%	8.7%	0.0%	0.0%	3.4%	3.5%	0.0%	9.1%	7.9%	0.0%	0.0%	3.5%	2.2%	13 GF	3.1%
14	PMC Paraphrase Multiple choice	1.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	0.2%	14 M	3.0%
15	PTGF Partial translation Gap filling	46.3%	14.9%	17.5%	22.4%	30.4%	15.6%	35.2%	27.1%	19.3%	20.8%	13.6%	13.2%	11.3%	26.3%	24.6%	22.4%	15 CV	2.2%
16	SA Short answer	7.4%	11.9%	10.5%	5.2%	2.2%	0.0%	2.8%	8.5%	8.8%	11.3%	6.1%	5.3%	4.8%	15.8%	1.8%	6.7%	16 PGF	2.2%
17	T Translation	5.6%	0.0%	1.8%	0.0%	2.2%	24.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	19.4%	12.3%	1.8%	5.0%	17 P	0.9%
18	T1 Translation (from English to Japanese)	14.8%	1.5%	1.8%	5.2%	13.0%	2.6%	5.6%	1.7%	3.5%	5.7%	7.6%	0.0%	11.3%	14.0%	21.1%	7.2%	18 PR	0.7%
19	T2 Translation (from Japanese to English)	7.4%	6.0%	0.0%	0.0%	3.9%	0.0%	7.0%	1.7%	1.8%	5.7%	0.0%	7.9%	0.0%	1.8%	10.5%	3.5%	19 BGF&CV	0.6%
20	OTHER Other type	0.0%	3.0%	1.8%	0.0%	0.0%	6.5%	25.4%	5.1%	5.3%	1.9%	4.5%	0.0%	0.0%	0.0%	0.0%	4.1%	20 C	0.2%
21	BGF&CV BGF+CV	0.0%	0.0%	8.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.6%	21 PMC	0.2%
22	MC&P MC+P	0.0%	0.0%	1.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	22 IT	0.1%
23	T/F True or False	0.0%	0.0%	10.5%	8.6%	0.0%	0.0%	11.3%	8.5%	8.8%	9.4%	7.6%	0.0%	0.0%	0.0%	7.0%	4.9%	23 MC&P	0.1%

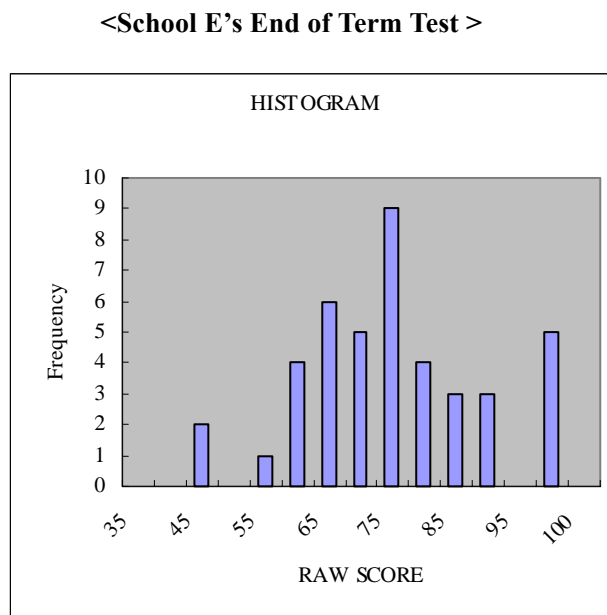
**Table 1.2 <Which item was used most?>**

### 3.3 Item analysis

School E's "English II" end of term test was chosen for item analysis. Compared to other schools' test, the style and item types of School E are almost the same and can be regarded as a standard type. School E is a coed senior high school and the students' English proficiency level is intermediate. The data was based on 42 samples that were randomly picked up from 197 juniors.

MEAN	72.7857
SEE	2.19691
MED	73.5
MODE	75
SD	14.2376
VAR	202.709
KURT	-0.1269
SKEW	0.09882
RANGE	59
MAX	41
MIN	100
SUM	3057

**Table 2.1**



Score	Frequency
35	0
40	0
45	2
50	0
55	1
60	4
65	6
70	5
75	9
80	4
85	3
90	3
95	0
100	5

**Table 2.2**

**Figure 1**

The test was consisted of fifty-seven items. Twenty-four out of fifty-seven questions were exactly copied from the exercises or quizzes that were on students' textbook and workbook. Thirty-eight out of fifty-seven questions (Vocabulary questions-7, Grammar questions-8, Translation questions-14, Reading comprehension questions-9) are based on two short stories (241 words and 253 words) of the textbook. This test was conducted within 50 minutes and the full mark was 100. Table 2.2, 2.1 and Figure 1 shows the result of this test.

In School E, students are supposed to be distributed official "Test News" a week before the test date. Teachers make this news to encourage students' studying. The paper provides information about a testing range and studying points such as "Text book. Page 93-103" and "Remember all new word and idioms. Mark off for spelling." And teachers also announce that test is consist of materials the student has learned in classes or homework and all items are based on them. From this aspect, the content validity of School E's test was not low because the testing point matched the teaching

point.

### 3.4 Reliability

To examine a reliability of this test, Cronbach's coefficient alpha was used. Allotments of each four categories are; Translation 48, Grammar 25, Vocabulary 7, and Reading Comprehension 20 and 100 stands for the full of this test. As Table 3 shows, the reliability of the whole test was 0.86. According to this result, it is reasonable to say this test is credible in terms of reliability.

Translation		Grammar		Vocabulary		Reading Comp		Pearson's correlation coefficient				
MEAN	33.095	MEAN	6.405	MEAN	2.476	MEAN	16.1		t	g	v	rc
SEE	1.2682	SEE	0.419	SEE	0.232	SEE	0.33	t	1.000			
MED	33	MED	6.5	MED	3	MED	16	g	0.685	1.000		
MODE	28	MODE	5	MODE	4	MODE	16	v	0.733	0.508	1.000	
SD	8.219	SD	2.715	SD	1.502	SD	2.139	rc	0.471	0.452	0.517	1.000
VAR	67.552	VAR	7.369	VAR	2.256	VAR	4.576	Cron	0.83	0.59	0.79	0.29
KURT	-0.096	KURT	-0.78	KURT	-1.19	KURT	0.32					
SKEW	-0.097	SKEW	0.271	SKEW	-0.52	SKEW	-0.22					
RANGE	34	RANGE	10	RANGE	4	RANGE	10					
MAX	14	MAX	1	MAX	0	MAX	10					
MIN	48	MIN	11	MIN	4	MIN	20					
SUM	1390	SUM	269	SUM	104	SUM	676					
Cronbach's coefficient alpha	0.83	Cronbach's coefficient alpha	0.59	Cronbach's coefficient alpha	0.79	Cronbach's coefficient alpha	0.29					

**$\alpha = 0.86205$**

Table 3

### 3.5 Item difficulty and Item discrimination power

Q No	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ID. Total *1	1.00	0.86	0.93	0.81	0.74	0.69	0.90	0.74	0.76	0.90	0.98	1.00	0.86	0.79	0.74
ID. Upper *2	1.00	1.00	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	1.00
ID. Lower *3	1.00	0.86	0.93	0.81	0.74	0.69	0.90	0.74	0.76	0.90	0.98	1.00	0.86	0.79	0.74
IDP. *4	0.00	0.33	0.93	0.50	0.25	0.67	0.25	0.67	0.58	0.25	0.00	0.00	0.25	0.33	0.42
Judgment	D	B	C	A	C	A	C	A	A	C	D	D	C	B	A

\*1 Total Item Difficulty    \*2 Item Difficulty of the upper level students

\*3 Item Difficulty of the lower level students    \*4 Item Discrimination Power

16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
0.86	0.90	0.93	0.71	0.69	0.81	0.69	0.98	0.26	0.60	0.93	0.33	0.26	0.48	0.36	0.95	0.93
1.00	1.00	1.00	1.00	0.91	1.00	0.91	1.00	0.73	0.91	1.00	0.91	0.64	0.55	0.73	1.00	1.00
0.86	0.90	0.93	0.71	0.69	0.81	0.69	0.98	0.26	0.60	0.93	0.33	0.26	0.48	0.36	0.95	0.93
<b>0.25</b>	<b>0.17</b>	<b>0.25</b>	<b>0.42</b>	<b>0.33</b>	<b>0.25</b>	<b>0.41</b>	<b>0.08</b>	<b>0.73</b>	<b>0.49</b>	<b>0.25</b>	<b>0.91</b>	<b>0.55</b>	<b>0.38</b>	<b>0.64</b>	<b>0.08</b>	<b>0.25</b>
C	<b>D</b>	C	A	B	C	A	<b>D</b>	A	A	C	A	A	B	A	<b>D</b>	C
33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
0.57	0.26	0.60	0.88	0.50	0.55	0.69	0.74	0.76	0.74	0.38	0.14	0.31	0.40	0.74	0.45	0.45
0.91	0.45	0.91	0.82	0.91	1.00	1.00	1.00	1.00	0.91	0.64	0.45	0.55	0.73	0.91	0.82	0.73
0.57	0.26	0.60	0.88	0.50	0.55	0.69	0.74	0.76	0.74	0.38	0.14	0.31	0.40	0.74	0.45	0.45
<b>0.74</b>	<b>0.45</b>	<b>0.66</b>	<b>-0.10</b>	<b>0.83</b>	<b>0.92</b>	<b>0.83</b>	<b>0.58</b>	<b>0.42</b>	<b>0.41</b>	<b>0.55</b>	<b>0.37</b>	<b>0.38</b>	<b>0.48</b>	<b>0.49</b>	<b>0.73</b>	<b>0.39</b>
A	A	A	<b>D</b>	A	A	A	A	A	A	A	B	B	A	A	A	B
50	51	52	53	54	55	56	57									
0.93	0.62	0.83	0.74	0.86	0.69	0.43	0.90									
1.00	0.73	0.82	0.91	1.00	1.00	0.82	1.00									
0.93	0.62	0.83	0.74	0.86	0.69	0.43	0.90									
<b>0.08</b>	<b>0.14</b>	<b>0.07</b>	<b>0.16</b>	<b>0.50</b>	<b>0.67</b>	<b>0.65</b>	<b>0.17</b>									
<b>D</b>	<b>D</b>	<b>D</b>	<b>D</b>	A	A	A	<b>D</b>									

Table 4

As Table 4 shows, every item difficulty and discrimination power were examined using Otomo's theory (1996). Judgment A-D was based on Ebel (1979). "A" means that the item is good to use, "B" means item is not bad but needs some modifications, "C" means the item is recommended to be modified, and "D" means that item is poor and should be removed or modified.

The first question we have to ask is why some items were judged "D". In this case, in my understanding, the reason why lower level students could answer these items correctly as well as upper is that some of these "D" judged items were exactly copied from student's textbook and workbook. In fact, ten out of twelve "D" judged items gave evidence to this supposition. According to this, it is reasonable to suppose that the copied items lowered item discrimination power because students had learned the questions and most students remembered the answers well.

The second question is how to deal with "D" judged items. Taken in the light of Ebel's judgment, twelve "D" judged items should be removed from this test. On the contrary, Shizuka (2003) pointed out that discriminating students' language ability is not the first purpose of the term test. In other words, item discrimination power is not important for term tests. Moreover, from viewpoints of the teachers, items such as easy



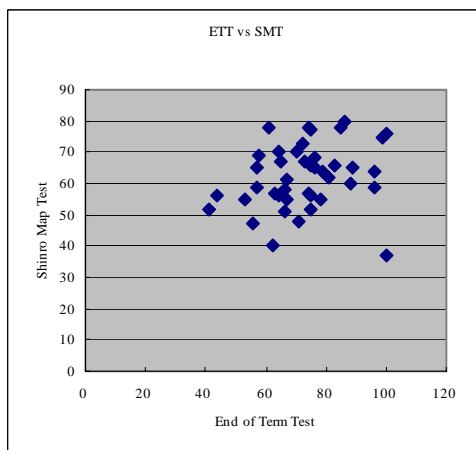
to answer are indispensable because many teachers try to avoid students failing completely. Therefore, they have to write these items even though they knew such items have lower discrimination power.

To take an example of School E, the purpose of this term test was to see if students understood the contents they learned in classes. Only term test score decide students' grade and grade are given using the theories of norm-referenced assessment. That is to say, no criterion is established for grading. Accordingly, from a point of view of testing theories, School E's test should be designed to discriminate students' abilities, that is to say, item discrimination power should be regarded as an important aspect. In spite of this, from a school educational view points, teachers are not willing to remove items which are easy to answer and item discrimination power is low.

The purpose of the term test is firmly related to the syllabus as well as the curriculum. And the way of evaluation changes according to the purpose. It would be difficult to decide which ideas is the best, however, on the theoretical side, Ebel's idea seems to be reasonable for School E's case, since grades are decided by norm-reference assessment theory. For this reason, "D" judged items should be removed or modified. On the other hand, Shizuka's idea looks really attractive. He also proposes that if all candidates had have acquired the contents of classes for a term test and taken the full mark, it would be the best condition for the term tests. However, to apply this ides, teachers would be required to reconstruct the whole evaluation system, syllabus and curriculum of their school. The information derived from his proposal is seen to be of value to a further research.

### **3.6 Correlation between End of term test and a Proficiency test**

What does the score of term test stand for? Did School E's term test measure English proficiency? To answer this question, I examined the correlation between the data of a term test with that of a proficiency test. A data of "Shinro Map Test" (SMT) created by Bessesse Corporation was used as a proficiency test. Figure 2 is the result of the correlation and Table 5 and Figure 3 shows the data of SMT.



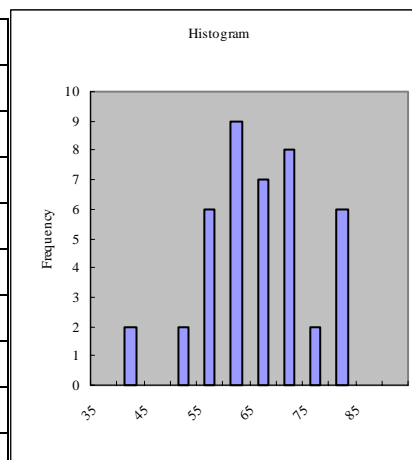
Correlation Coefficient = 0.249695

**Figure 2**

**<SMT>**

MEAN	62.2381
SEE	1.58492
MED	63
MODE	78
SD	10.2715
VAR	105.503
KURT	-0.1426
SKEW	-0.261
RANGE	43
MAX	37
MIN	80
SUM	2614

**Table 5**



**Figure 3**

As Figure 2 shows, correlation between the term test and SMT is negative. It is obvious that School E's term test has no correlation with a proficiency test.

Term test is usually categorized as a progress achievement test and it is said that the purpose of the progress achievement test is to evaluate the achievement level of criterions established by teachers. From Figure 2, School E's end of term test did not measure student proficiency. However, it would be difficult to say that the result of this test means a student's achievement level of criterions because it was not a clearly available criterion established for this test by teachers of School E. According to my interview with one of a teacher of School E, it seems that criterion of assessment and syllabus are still not regarded as important factors for assessing and evaluating students. They make term tests and mark them just to discriminate the student. For these reasons, teachers cannot provide appropriate information and advice to their students based on the test score. In other words, teachers cannot explain what each student's raw score means properly. In my understanding, proper syllabus designing and criterion establishment would clear eradicate problems.

#### 4. Conclusion

The analysis employed in this paper revealed substantial evidence for the problem of Japanese term tests relating to the differences between the classroom activities and government policy. Moreover, establishing a firm criterion for assessing and constructing a proper syllabus of each school is indispensable in order to enhance students English studying. Further investigation is necessary to improve the testing of term test in Japan.

## References

- Alderson, J. Charles. Clapham, Caroline. and Wall, Dianne. (1995). *Language Test Construction and Evaluation*. Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L. F. (1996). *Language Testing in Practice*. Oxford University Press. (1999).
- Ebel, R.L. (1979). *Essentials of Educational Measurement*. (Third ed.). Prentice-Hall, Inc.
- Hughes, Arthur. (2003). *Testing for Language Teachers*. (Second ed.). Cambridge University Press.
- Kiyokawa, Hideo. (1990). *Eigo Kyouiku Nyuumon*. Taishukan.
- Matsuzawa, Shinji. Sano, Masayuki. and Yoneyama, Asaji. (2002). *Eigo Kyoushi no Tame no Atarashii Hyoukaihou*. Taishukan
- Minoru, Wada. (trans.) (1999). *Gengo Testo no Kisochishiki. Testing in Language Programs*. By Brown, J.D. Taishukan.
- Otomo, Kenji . (1996). *Koumoku Outou Riron Nyuumon*. Taishukan.
- Rechards, C. Jack. & Schmidt, Richard. (2002). *Dictionary of Language Teaching & Applied Linguistics*. Longman.
- Shizuka, Tetsuhito. (2002). *Eigo Tesuto sakusei no Tatsujin Manyualu*. Taishukan.
- Wakabayashi, Shunsuke. Negishi, Masashi. (1993). *Musekininn na Tesuto ga Ochikobore wo Tsukuru*. Taishukan.

## Appendix

### 1 <Accent>

Q: Choose the number corresponding to the syllable which contains the primary stress.

1 math-e-mat-ics    2 re-la-tion-ship    3 ed-u-cate    4 in-ter-pret-er    5 hon-or-able  
1 2 3 4    1 2 3 4    1 2 3    1 2 3 4    1 2 3

### 2 <Correction>

Q: Correct two errors in the sentence (8).

... (8) Maybe you know much about it than we are ! ...

### 3 <Conversion>

Q: Put the following sentences in negative.

(1) My brother is in the room.    (2) You are a good girl.

### 4 <Banked-Gap filling>

Q: Fill in the blanks with (*am, are, is*).

(1) This ( ) my umbrella.    (2) They ( ) my brothers.    (3) I ( ) John Smith.

### 5 <Correction Gap filling>

Q: Correct the word inside ( ) to complete sentence.

(1) I sat ( ) at the beautiful sunrise. (look)    (2) He remains ( ). (seat)

6 <Gap filling>

Q: Fill in the blanks with an appropriate word.

(1) A: What's in the latest magazine, Ken?" B: I don't know. I ( ) not ( ) it yet.

7 <Information transfer>

Q: After reading a story, illustrate on your answer sheet (1) the place and (2) the environment where "Jeff" lived.

8 <Matching>

Q: Match A-F on the left with 1-8 on the right to make each sentence correct.

A: Hoshino Michio

1: is minus 40 degrees Celsius.

B: The mayor of Shishmarf

2: lived Alaska and took pictures.

C: -----

3: -----

9 <Multiple choice>

Q: Choose the best answer from a-c.

(1) George and I (a: listen b: listens c: listened ) to radio everyday.

10 <Ordering Task>

Q: Put the following words in the correct order.

(1) (about / worried / the test / I'm) next month.

11 <Paraphrase>

Q: Add missing words into the underlined sentence (7) to make a correct English sentence.

... Of course I know it, but (7) not much about it. Do I write ...

12 <Pronunciation>

Q: Choose the word where the underlined part is pronounced differently.

(1) 1: meal 2: indeed 3: ready 4: season

13 <Paraphrase Gap filling>

Q: Fill in the blanks in appropriate word and make the sentence 1 and 2 mean the same.

(1) 1, It was cold this morning. It is still cold now. 2, It ( ) ( ) cold ( ) this morning.

14 <Paraphrase Multiple choice>

... Foreign lands were so far away for us in those days; (6) how could a girl ever hope to make it home safely? ...

Q: Choose the number corresponding to the sentence closest in meaning to (6).

a, it was very difficult for a girl to travel abroad and get back home safely.

b, girls ...

15 <Partial translation Gap filling>

Q: Fill in the blanks to make each sentence match Japanese meaning.

(1) (A Japanese translation comes here.) The man ( ) over the fence.

(A) It's (a ), isn't it? (A Japanese meaning of the blank part comes here.)

16 <Short answer>

Q: After reading the passage, answer the following questions.

(1) What was the new born baby named? (2) Did mother's voice sound happy?

17 <Translation> (Word translation)

18<Translation 1> (English to Japanese)

Q: Translate the following sentences in English into Japanese.

(1) English is certainly a very useful language for global communication, but it is, after all, just one of those many languages.

19<Translation 2> (Japanese to English)

Q: Translate the following sentences in English.

(1) (A Japanese sentence comes here.)      (2) (A Japanese sentence comes here.)

20<Other>

-----sentence-----.1-----sentence-----.3-

-----

sentence-----.4-----sentence-----.5-----sentence -----

Q: Select the best position from 1-5 to add the following sentence to complete a story.

But our appetite sees to it that we don't satisfy our hunger.

21<Banked-Gap filling + Conversion>

22<Multiple choice + Paraphrase>

23<True or False>