High School English Textbooks and College Entrance Examinations: A Comparison of Reading Passage Difficulty

Shinji Kimura
Kwansei Gakuin University
Brad Visgatis
Osaka International College for Women

This paper examines the difficulty level of 48 entrance examination reading passages, taken from tests at 33 junior colleges, and compares them with the difficulty level of 66 passages taken from 4 approved high school English textbooks using readability indices. Though wide variation in readability scores was recorded, overall results indicate test reading passage difficulty to be significantly higher than the difficulty of textbook reading passages. A serious mismatch between official test candidate requirements and what is truly required of the applicants may exist. Schools are encouraged to review their literature for prospective students and/or adapt their tests to help ensure fairness and validity.

この論文は、33の短期大学の入学試験からとられた48の読解文と、4種の文部省検定教科書からとられた6 6 の文章の難易度を、リーダビリティ指標を使って比較をしたものである。リーダビリティのスコアには幅があったが、全般的な結果としては、試験の文章の難易度は、教科書の文章の難易度に比べて、著しく高かった。公式の受験資格と、実際に受験者に期待される学力の間には深刻なギャップがあるのかもしれない。公正かつ妥当な試験を実施するためには、学生募集要項を改訂するか、試験を教科書に合わせるかすることが望まれる。

The importance of college and university entrance examinations in Japan is well known. All parties involved in the examination process (high school students; teachers and administrators; college teachers, test developers and administrators; parents and relatives) devote considerable resources to them.

Given this importance, it is essential that the examination process maintain the highest standards of quality. Or, in other words, "the more important the decision to be made, the greater the effort that should be expended

in assuring that [a] test is reliable and valid" (Bachman, 1990, p. 56).

One way to insure such a high standard is through a regular process of review and evaluation. A number of authors discuss the issue of evaluating test quality (Alderson, Clapham & Wall, 1995; Bachman, 1990; Brown, 1995; Henning, 1987) and highlight the types of evaluation possible, including estimating reliability and assessing different types of validity. Other aspects of the testing process, such as comparing test development practices with accepted theory, are also possible.

Unfortunately, most of these types of evaluation and review require access to test results. In Japan, though copies of tests are regularly published, test results themselves are held in confidence and access is difficult to obtain. Without access to test results, calculating reliability coefficients and determining validity is very difficult.

However, a few methods of evaluation do exist which do not require detailed access to test results. One method is to take the published test questions, readminister them to a new group of subjects, and then compare the subjects' results with their results on other types of tests. One study which used this approach looked at the validity of written tests of pronunciation (Buck, 1989) and found scores on such tests had no significant relationship to productive tests of pronunciation ability.

Another approach involves examining the characteristics of published tests and analyzing them from a theoretical standpoint, using modern test theory to highlight areas which appear to be substandard. One example of such an investigation was the study carried out by Brown and Yamashita (1995) where they explored various test aspects, including item type, difficulty of reading passages, differences between public and private institutions and types of skills measured on the tests. Here, they used accepted standards of testing theory to point out areas where there was a need for improvement.

Another type of evaluation involves content validation. According to Alderson, Clapham and Wall,

> content validation involves 'experts' making judgements in some systematic way. A common way for them is to analyse the content of a test and to compare it with a statement of what the content ought to be. Such a statement may be the test's specifications, it may be a formal teaching syllabus or curriculum, or it may be a domain specification. (1995, p. 173)

As such, content validation is one way researchers outside of the testing process can approach the evaluation of a test.

One example of this would be to compare the difficulty of test materials with that stated in the test's specifications and to determine

whether the test was set at an appropriate level for the targeted examinee. Difficulty is seen as important by a number of researchers. Henning (1987) finds that "the single most important characteristic of an item to be accurately determined is the difficulty" (p. 49), and that "when tests are rejected as unreliable measures for a given sample of examinees, it is due not so much to the carelessness of the item writers as to the misfit of item difficulty to person ability" (op. cit.).

Tests which are at a level of difficulty inappropriate for the targeted audience are compromised. First, such tests display a skewed distribution of scores which reduces the test's reliability (Bachman, 1990; Henning, 1987). Or, as Henning states, "Tests that are too difficult or too easy for a given group of examinees often show low reliability" (1987, p. 49). Second, "[i]f the test is too easy or too difficult for a particular group, this will generally result in a restricted range of scores or very little variance" (Bachman, 1990, p. 220). Bachman (1990) goes on to argue that a test which contained

> tasks at levels of difficulty that are inappropriate for the ability level of the group being tested . . . [or] . . . with all items at the same level of difficulty would not be a very accurate measure for individuals whose abilities are either greatly above or greatly below that level. Likewise, neither extremely easy nor extremely difficult items will provide very accurate measures for a group of individuals of relatively homogeneous intermediate ability. (p. 36)

Clearly, the assessment of test difficulty with regard to the targeted level of difficulty can yield important information for evaluating and ameliorating the entrance examination system in Japan.

Purpose

The purpose of this study is to evaluate the difficulty of Japanese junior college English entrance examination reading passages and compare that with the targeted difficulty level as stated in the test information given out to the applicants.

Reading passages were chosen because of their widespread use on language examinations and because of their perceived importance in assessing foreign language ability. For example, a poll by the Japan Association of College English Teachers (JACET) found that 96.8% of the respondents cited reading as a domain covered in Entrance Examinations, and of those respondents 73% percent gave it a weight of between 50% and 80% of the total test points (Tajima, 1993).

Our research question is this: *Is there a significant difference between the difficulty levels of passages on college English entrance examinations and the stated target level of passage difficulty?*

Method

Establishing the target level of difficulty

First, we wanted to establish the targeted level of the reading passages on the examinations. We argue that the difficulty levels correspond closely to the difficulty level of materials used in high school courses because of the following:

1. The Ministry of Education issues guidelines to colleges and universities indicating how the selection process for incoming students should be carried out (Ministry of Education, 1993) and expects those schools to set tests accordingly. Though the Ministry of Education does not require schools to state the exact level of the tests that they administer, schools are expected to make reference to the particular high school course of study the perspective applicants should have completed. Though recent changes have been implemented in the high school curriculum[1], current college and university students studied most, if not all three of the following English reading courses offered in high school: *Eigo I, Eigo II,* and *Eigo IIb,* with *Eigo I* being the most basic and *Eigo IIb* the most advanced. Thus, colleges and universities when setting their tests officially stated whether they were intended for students who had completed *Eigo I, Eigo II,* or *Eigo IIb.*
2. Information given out by colleges and universities makes reference to the particular high school course of study the prospective applicants should have completed. Examples of this can be found in the promotional literature issued by individual schools as well as by examination of some of the common test preparation guidebooks widely available.

Developing the databases of reading materials

Next we developed two databases of reading passages. Before collecting passages, we decided to limit our investigation to the *Eigo II* level of materials. This was done for two reasons. First, it was the level most commonly used by our department and would provide the most useful information for our own purposes. Second, according to figures in Kimura and Visgatis (1992), this level appeared to be the most com-

monly targeted one among junior colleges, with 89 of 146 schools setting it as their testing level.

One database was made up of passages taken from four high school reading textbooks: *Creative English II* (Kakita et al., 1992), *Mainstream II* (Ando et al., 1991), *Raccoon II* (Onodera et al., 1992.) and *Enjoy English II* (Hasegawa, Ishii, Hayakawa, Yamaguchi, & O'Conner, 1992). All of these textbooks were in use during the 1992 school year. They, or more recent editions, are currently still in use for second and third year high school students. In all, 66 textbook reading passages were selected for this database. To develop this database, passages were electronically scanned and converted into computer text files with the use of optical character recognition software. In all, 66 passages were selected.

The second database was made up of reading passages taken from sample entrance examinations. These passages were taken from the examination guidebook *Zenkoku Tanki Daigaku Nyuushi Mondai Seikai, Eigo•Kokugo* [All-Japan Junior College Entrance Examination Problem Solutions, English and Japanese] (*Zenkoku*, 1992). This guidebook contained information on tests given at 74 two-year women's colleges throughout Japan. Two-year women's colleges were chosen because they corresponded best with the level of students accepted by our own institution and were accordingly the level with which we were most familiar. From this guidebook we chose reading passages from tests offered at 33 different colleges. All of them were for students graduating from the *Eigo II* course of high school study. All of the tests were administered by their respective schools during the 1992 entrance examination period.

In all, 48 reading passages were selected and typed into a Macintosh computer using word processing software. For reading passages containing blank spaces, the appropriate word or words to complete the item were inserted before the readability scores were generated. Additionally, no distinction was made between passages which contained glossed items and those which did not. (This issue is addressed more fully in the Discussion.)

Analysis

Reading passages in both of the databases were evaluated for readability using *CorrectGrammar 3.0* (Writing Tools Group, Inc., 1992) for the Macintosh computer. This software package contained three readability measures: the Flesch Reading Ease (FRE), Gunning's Fog Index (GF) and the Flesch-Kincaid Reading Grade Level (FK). These three

formulas are measures which estimate the difficulty level of the reading passages by evaluating such textual features as the number of syllables per word, the average number of words per sentence, number of sentences per paragraph, etc.

The Flesch Reading Ease scale ranges from 0 to 100, with 100 indicating the easiest to read. The Flesch-Kincaid and Fog readability indexes are expressed in grade levels, normalized on the American educational system and indicating the appropriate grade level for the reader.

These three indices were chosen for a number of reasons: they have been adapted for computers which enable computerized checking; they are widely available and are often bundled with mainstream word processing packages; they have been in use for over twenty years and are widely known. As Klare points out, "well over 1,000 readability references can be found in the library" (1984, p. 682). Many of these involve reference to the three measures used here. In addition, previous research on entrance examination questions (Brown and Yamashita, 1995) has used similar readability indices. (For more information on and discussion of readability formulas, see Harrison, 1980; Klare, 1984; Writing Tools Group, Inc., 1992; Zakaluk and Samuels, 1988.)

Results

The readability statistics of passages taken from the entrance examinations are given in Table 1. They show a wide range of levels on all three readability measures. The Flesch Reading Ease scores ranged from 41.900 to 92.900, with a mean of 64.804. Gunning's Fog Index scored from a low of 3.800 to a high of 17.300, with a mean of 10.902. The Flesch-Kincaid showed scores ranging from a minimum of 1.600 (roughly equivalent to just under U.S. 2nd grade elementary school level) to a maximum of 13.200 (roughly equivalent to the U.S. sophomore level in college), with a mean of 8.252. The standard deviations were respectively, 13.243, 2.946 and 2.715.

Readability statistics for passages taken from the high school textbooks are given in Table 2. They, too, evince a high degree of variation, with Flesch Reading Ease scores ranging from 46.000 to 98.300 and a mean of 75.985, the Gunning's Fog Index scores ranging from 3.600 to 13.500 and a mean of 8.326, and Flesch-Kincaid scores ranging from 1.300 to 11.800 and a mean of 5.985. Standard deviations were respectively, 10.829, 2.221 and 2.208.

Discussion

As a whole, the readability statistics for both the examination and textbook passages show surprisingly wide variation. In some cases, this variation measures up to approximately 11 U.S. grade levels on the

Table 1: Test Passage Readability Statistics

Passage	FRE	GF	FK	Passage	FRE	GF	FK
1	67.7	8.9	6.4	25	69.3	10.2	6.8
2	60.7	10.8	8.2	26	73.5	10.1	7.2
3	72.7	9.7	6.9	27	45.4	14.4	12.7
4	44.1	15.5	12.0	28	63.9	10.4	7.4
5	45.1	15.6	12.2	29	59.0	13.3	11.1
6	68.6	9.8	7.2	30	73.3	9.3	6.1
7	70.7	9.1	6.7	31	68.8	9.7	7.9
8	92.9	3.8	1.6	32	82.0	8.5	6.4
9	58.4	11.2	7.9	33	74.3	7.7	5.6
10	41.9	17.3	13.2	34	92.2	4.3	2.2
11	82.9	7.0	4.9	35	52.9	12.6	10.6
12	87.1	7.5	4.5	36	59.4	11.8	8.2
13	80.7	8.1	5.0	37	53.9	16.2	12.6
14	58.5	12.9	10.1	38	56.7	12.8	10.4
15	76.7	8.3	5.9	39	60.0	12.3	10.4
16	52.0	12.1	9.5	40	65.2	10.2	7.2
17	76.2	9.5	7.3	41	55.5	12.6	10.0
18	62.1	12.7	9.1	42	76.3	11.9	9.0
19	57.0	11.0	8.8	43	85.9	5.6	3.5
20	66.2	11.0	8.6	44	52.3	12.4	10.2
21	59.5	9.8	8.6	45	47.4	14.8	10.7
22	51.5	13.9	12.1	46	54.4	12.3	9.6
23	77.6	9.1	6.3	47	76.3	8.6	6.5
24	44.4	15.4	11.7	48	57.5	11.3	9.1

	FRE	GF	FK
Mean	64.804	10.902	8.252
Maximum	92.900	17.300	13.200
Minimum	41.900	3.800	1.600
Range	51.000	13.500	11.600
Standard Deviation	13.243	2.946	2.715

FRE = Flesch Reading Ease, FK = Flesch-Kincaid Level, GF = Gunning's FOG Index

Flesch-Kincaid, from low elementary school up to the sophomore level of college. Gunning's Fog Index and Flesch Reading Ease showed similar patterns of variation.

One consistent trend, however, is that examination reading passages are more difficult, often by a factor of a few Flesch-Kincaid grade levels, than high school textbook passages. For example, the mean Flesch-Kincaid score of entrance examinations was 8.252, while that of the high school

Table 2: Textbook Passage Readability Statistics

Passage	FRE	GF	FK
1	71.8	9.0	6.8
2	60.9	12.3	9.6
3	73.4	8.5	6.0
4	70.2	8.6	6.9
5	66.5	8.0	7.6
6	79.6	7.7	5.3
7	87.2	7.2	4.6
8	75.8	9.5	6.3
9	88.0	6.6	4.0
10	68.0	10.4	7.5
11	73.1	9.5	6.7
12	57.6	11.3	8.8
13	90.6	5.2	2.8
14	60.4	12.9	9.9
15	59.0	10.9	8.3
16	77.6	9.3	6.4
17	88.5	5.7	3.6
18	85.1	6.0	3.9
19	76.2	8.8	6.6
20	72.5	9.2	6.9
21	78.5	7.7	5.5
22	87.7	6.4	3.6
23	75.0	7.8	5.7
24	67.7	9.3	7.6
25	77.7	8.7	6.0
26	71.6	9.3	7.1
27	46.0	13.5	11.8
28	95.7	3.9	1.5
29	71.5	9.8	7.5
30	77.0	8.4	6.4
31	74.6	9.3	7.1
32	85.0	6.7	4.5
33	59.2	11.0	9.1
34	96.3	3.6	1.6
35	98.3	3.7	1.3
36	82.6	7.7	4.7
37	82.1	6.8	4.8
38	73.5	9.4	6.7
39	93.8	4.6	2.2
40	65.1	10.9	9.0
41	78.6	7.9	5.3
42	65.9	9.7	7.6
43	81.7	7.6	5.0
44	85.9	5.2	3.2
45	70.7	9.3	6.5
46	84.6	5.4	3.3
47	83.9	7.2	4.4
48	65.0	11.5	8.0
49	87.2	7.2	4.2
50	79.0	8.0	5.7
51	76.6	7.1	5.2
52	86.4	6.4	4.2
53	71.8	8.8	5.9
54	71.9	10.6	7.9
55	58.0	11.0	8.7
56	68.1	9.1	7.4
57	78.7	8.1	5.0
58	71.7	10.4	7.7
59	78.3	8.2	5.9
60	59.1	10.0	9.0
61	75.6	8.1	6.1
62	94.5	4.1	2.2
63	72.5	9.3	6.5
64	67.7	9.7	7.7
65	69.3	10.0	7.8
66	91.4	4.5	2.4

	FRE	GF	FK
Mean	75.985	8.326	5.985
Maximum	98.300	13.500	11.800
Minimum	46.000	3.600	1.300
Range	52.300	9.900	10.500
Standard Deviation	10.829	2.221	2.208

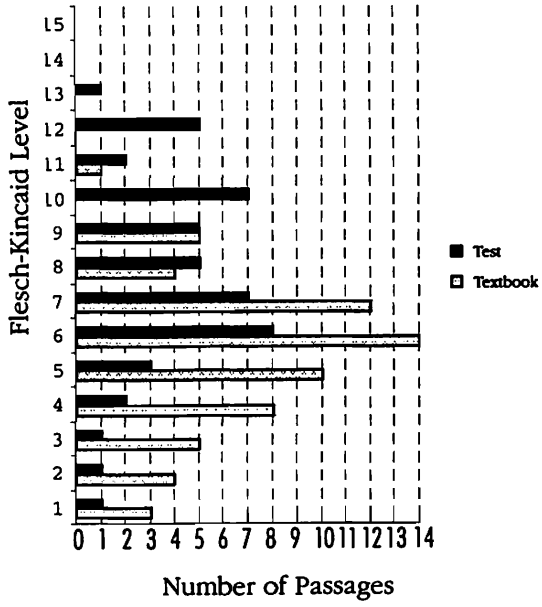
FRE = Flesch Reading Ease, FK = Flesch-Kincaid Level, GF = Gunning's FOG Index

reading passages only measured 5.985. This amounts to more than two U.S. grade levels of difference. Both the Gunning's Fog Index and Flesch Reading Ease showed similar patterns, with a difference of more than 2-points in the former, and more than 9-points in the latter.

These differences were statistically significant. The MANOVA results indicated overall multivariate significance at $p < .001$ (for three multivariate statistics: Pillais, Hotellings, and Wilks). Thus, univariate ANOVA comparisons for each dependent variable were justified. Each of these comparisons also turned out to be significant at $p < .001$ (F for FRE = 24.592; F for GF = 26.946; and F for FK = 22.548). These results indicate that there is only a one in 1,000 chance that the mean differences observed here were due to chance alone. Some more examples are telling: While there are 15 examination passages with Flesch-Kincaid reading scores at 10 or above, there is only one among the textbook passages (see Graph 1).

Given that, for example, the Flesch-Kincaid readability scores are designed to correlate roughly to U.S. grade levels, it can be argued that the difference in scores is significant: expecting students to be able to read

Graph 1: Flesch-Kincaid Reading Passage Distributions



materials three or more grade levels above the materials they have been exposed to challenges the credibility of the examination passages, and by association, that of the targeted level. This is even more striking after considering that students using textbooks are free to read the passages at home, consult reference works (i.e. dictionaries), and are not subjected to the rigorous time constraints found under examination conditions.

However condemning these statistics may seem, there are several points to consider when interpreting the results.

The first consideration is that the college entrance examinations, being designed to select the above-average members of the high school cohort (i.e. those who best deserve admittance to tertiary education), need to be set at a level above what the average high school student would be expected to cope with. However, there remains the critical question of just how high the target level for Japanese students needs to be. The difference of over two Flesch-Kincaid reading grade levels found between the most difficult test reading passage and the most difficult textbook reading passage may be so great as to seriously compromise that test's reliability, by forcing students into guessing at answers rather than using their comprehension of the passage. More research is needed to evaluate this.

Second, it might be assumed that students are faced with progressively more difficult reading materials as they proceed through the high school curriculum, thus being amply prepared for the difficult reading passages found on entrance examinations. Unfortunately, this is not borne out by the textbook materials. Examination of the difficulty patterns of textbook reading passages (see Table 3) shows that the highest average Flesch-Kincaid reading level does not appear in the last third of any of the textbooks, and only two of the textbooks have the most difficult Gunning-Fog result in the final third. If the chapters in the books are used sequentially, students will not be facing the most difficult passages at the end of their high school tenure.

Third, readability formulas have been criticized along a number of different lines. One line challenges the use of formulas normalized against native-speaker proficiencies with non-native readers (Carrell, 1987). As rebuttal, we can only argue that the converse, that is, that a null or negative relationship between readability scores and reading difficulty for non-native vs. native speakers is counterintuitive. Why should non-native speakers be expected to be able to read materials that native-speakers would likely find difficult? Indeed, this view is also supported by Alderson, Clapham and Wall, who encourage native-speaker trials for objective tests, as most test candidates "cannot be expected to pro-

Table 3: Difficulty Patterns of High School Reading Passages

Creative English II	Unit	Units 1-6	Units 7-11	Units 12-18	Difficulty Pattern		
	FRE	70.40	78.42	72.28	Hardest	Easiest	Medium
	GF	9.02	8.64	9.22	Medium	Easiest	Hardest
	FK	7.03	5.82	6.63	Hardest	Easiest	Medium
Mainstream English II	Unit	Units 1-6	Units 7-12	Units 13-18	Difficulty Pattern		
	FRE	79.17	71.70	81.73	Medium	Hardest	Easiest
	GF	7.65	9.08	7.12	Medium	Hardest	Easiest
	FK	5.37	6.92	5.00	Medium	Hardest	Easiest
Raccoon II	Unit	Units 1-5	Units 6-9	Units 10-14	Difficulty Pattern		
	FRE	79.42	78.03	78.28	Easiest	Hardest	Medium
	GF	7.88	7.60	8.12	Medium	Easiest	Hardest
	FK	5.48	5.28	5.28	Easiest	Medium	Medium
Enjoy English II	Unit	Units 1-6	Units 7-11	Units 12-18	Difficulty Pattern		
	FRE	73.95	71.18	78.50	Medium	Hardest	Easiest
	GF	8.65	9.16	7.62	Medium	Hardest	Easiest
	FK	6.27	7.00	5.45	Medium	Hardest	Easiest

FRE = Flesch Reading Ease, FK = Flesch-Kincaid Level, GF = Gunning's FOG Index

duce as high a level of language as well-educated native speakers . . . [and] any items which turn out to be too difficult for such native speakers should be omitted" (1995, p. 97).

Another line challenges the reduction of the determination of reading difficulty to analysis of textual features, such as number of words per sentence, percentage of multi-syllabic words, and so forth, without regard to other factors, such as motivation for reading and the influence of schematic knowledge (Carrell, 1987; Harrison, 1986).

To this, we would like to propose that, first, test developers are not likely to select passages which require extensive schematic knowledge to understand, and second, that the lack of student reading motivation does not necessarily become a factor. After all, the desire to enter the college or university of their choice should provide students with ample instrumental motivation for reading the passages. Assuming these two propositions are correct, the role of textual characteristics assumes a larger, if not commanding, role in dictating passage difficulty.

Fourth, it must be remembered that passage difficulty is not necessarily indicative of question difficulty. It is possible, and even likely, that some of the difficult reading passages are followed by relatively easy questions—questions which do not require a true understanding of the passage in order to answer successfully.

A related concern is the inclusion of passages with glossed items. Admittedly, the difficulty level of a passage is reduced if some of the harder

Table 4: Readability Patterns of Glossed and Unglossed Test Passages

Glossed Test Passages

Passage	FRE	GF	FK
1	67.7	8.9	6.4
9	58.4	11.2	7.9
11	82.9	7.0	4.9
12	87.1	7.5	4.5
14	58.5	12.9	10.1
16	52.0	12.1	9.5
19	57.0	11.0	8.8
20	66.2	11.0	8.6
25	69.3	10.2	6.8
26	73.5	10.1	7.2
29	59.0	13.3	11.1
39	60.0	12.3	10.4
41	55.5	12.6	10.0
46	54.4	12.3	9.6

Unglossed Test Passages

Passage	FRE	GF	FK
2	60.7	10.8	8.2
3	72.7	9.7	6.9
4	44.1	15.5	12.0
5	45.1	15.6	12.2
6	68.6	9.8	7.2
7	70.7	9.1	6.7
8	92.9	3.8	1.6
10	41.9	17.3	13.2
13	80.7	8.1	5.0
15	76.7	8.3	5.9
17	76.2	9.5	7.3
18	62.1	12.7	9.1
21	59.5	9.8	8.6
22	51.5	13.9	12.1
23	77.6	9.1	6.3
24	44.4	15.4	11.7
27	45.4	14.4	12.7

Passage	FRE	GF	FK
28	63.9	10.4	7.4
30	73.3	9.3	6.1
31	68.8	9.7	7.9
32	82.0	8.5	6.4
33	74.3	7.7	5.6
34	92.2	4.3	2.2
35	52.9	12.6	10.6
36	59.4	11.8	8.2
37	53.9	16.2	12.6
38	56.7	12.8	10.4
40	65.2	10.2	7.2
42	76.3	11.9	9.0
43	85.9	5.6	3.5
44	52.3	12.4	10.2
45	47.4	14.8	10.7
47	76.3	8.6	6.5
48	57.5	11.3	9.1

	Glossed				Unglossed		
	FRE	GF	FK		FRE	GF	FK
Mean	64.393	10.886	8.271	Mean	64.974	10.909	8.244
Maximum	87.100	13.300	11.100	Maximum	92.900	17.300	13.200
Minimum	52.000	7.000	4.500	Minimum	41.900	3.800	1.600
Range	35.100	6.300	6.600	Range	51.000	13.500	11.600
Std. Deviation	10.690	1.961	2.058	Std. Deviation	14.307	3.294	2.971

terms are explained in an easier to understand format, such as rewording in English or through translation into Japanese. In addition, glossing is not the only factor which may influence understanding. Occasionally, the way the questions for a passage are presented may give a helpful indication as to the meaning of the passage content.

To try to estimate the impact the inclusion of glossed items had on the readability statistics, we examined the glossed passages more carefully. In all, 14 of the 48 passages contained a total of 38 items (words or expressions) which were glossed. The total number of words in the glossed items was 45. This amounted to less than 1 percent of the 4,904 words found in those passages. In addition, we used the Mann-Whitney Test to compare the readability levels of glossed and non-glossed passages and found no

significant difference (FK, p>.999; FRE, p=.856; GF, p=.874).

Fifth, the reasoning involved in identifying the targeted level of proficiency by reference to schools' promotional materials may be suspect. Private colleges may base their targeting on the overall high school language curriculum, which includes a number of areas which are only vaguely defined by the Ministry of Education. One such area involves supplementary materials. These are used in addition to the course textbook and may be of a higher level of difficulty. Determining the true level of difficulty of all of the materials used at the high school level is more problematic given the lack of clear specification. This is one point where further research is needed.

Conclusion

Our research question was, *is there a significant difference between the difficulty levels of passages on college English entrance examinations and the stated target level of passage difficulty?*

The answer to this question is "yes." A consistent pattern emerges of examination passages set at a level significantly above that which is expected by the Ministry of Education and reflected in the reading materials found in the approved high school textbooks. In addition, these results for junior colleges are in line with the results found by Brown and Yamashita (1995) for passages taken from entrance examinations at prestigeous public and private universities. This indicates that the results here can be extrapolated to most institutions of higher education in Japan.

These results may be interpreted in a number of ways.

It is possible that those charged with preparing entrance examinations are not aware of the materials currently used in high school. Or, test developers may simply be lax in their materials vetting procedures. In either case, the solution to this would be development of better examination writing guidelines, improved test specifications and rationalized vetting procedures.

Another possibility is that the colleges (and universities) are only paying lip service to guidelines issued by the Ministry of Education regarding entrance examinations, and reference to a particular course of study in high school has no bearing upon the actual test material generated. If so, schools may be setting standards according to some other benchmark. *If this is the case, those schools should make that fact clear to students in advance.*

Finally, given the test development climate in Japan where piloting of examination questions is quite rare, the use of readability formulas to

assist in the process of selecting reading passages may prove useful. As one measure among several it can provide insights into the relative difficulty of various passages, enabling test-developers to make examinations with better reliability and validity.

Shinji Kimura is now a lecturer at Kwasei Gakuin University. His research interests include computer aided instruction and reading.

Brad Visgatis is an associate professor at Osaka International College for Women. His research interests include task-based teaching and curriculum design.

Note

1. For a discussion of the changes, see Wada and Koike (1990).

References

Ando, S., Hale, D., Kawazumi, T., Kubouchi, H. Komura, M., Chigusa, M., Nakano, M., Harada, T., Hibino, H., Mizuno, M., & Miyamoto, H. (1991). *Mainstream English II.* Tokyo: Zoushindou.

Brown, J. D. (1995). *Testing in language programs.* Upper Saddle River, NJ: Prentice-Hall Regents.

Brown, J. D., & Yamashita , S. O. (1995). English language entrance examinations at Japanese universities: What do we know about them? *JALT Journal, 17,* 7-30.

Buck, G. (1989). Written tests of pronunciation: Do they work? *ELT Journal, 43,* 50-56.

Carrell, P. L. (1987). Readability in ESL. *Reading in a foreign language, 4,* 21-40.

Harrison, C. (1980). *Readability in the classroom.* London: Cambridge University Press.

Harrison, C. (1986). New directions in text research and readability. In A. Cashdan (Ed). *Literacy: Teaching and learning language skills* (pp. 61-81). Oxford: Blackwell.

Hasegawa, K., Ishii, S., Hayakawa, K., Yamaguchi, K., & O'Conner, W. F. (1992). *Enjoy English II.* Tokyo: Kyouiku Shuppan.

Henning, G. (1987). A *guide to language testing: Development, evaluation, research.* Cambridge: Newbury House Publishers

Kakita, N., Nagata, T., Tanabe, K., Temma, M., Hatori, H., Okano, K., Mura, S., Minamimura, T., Hayashi, H., Tada, Y., Okihara, K. Koshino, T., Dan, I., Yasumori, Y. & Davenport, C. (1992). *Creative English II.* Tokyo: Daiichi Gakushusha.

Kimura, S., & Visgatis, B. (1992). University entrance examination and testing theory. *The Bulletin of Teikoku Gakuen, 18,* 167-177.

Klare, G. R. (1984). Readability. In P.D. Pearson (Ed.), *Handbook of reading*

research (pp. 681-738). New York. Longman.

Ministry of Education. (1993). *Monbushou koutou kyouiku kyokuchou tsutatsu - Daigaku nyuugakusha senbatsu jisshi youkou* [Notification by the director of the Bureau of Higher Education, Ministry of Education: Outline on university entrance examination candidate selection procedures]. Tokyo: Ministry of Education.

Onodera, T. , Lummis, C. D., Takahashi, S., Murayama, K. Ishida, S. Kurata, M., & Ikeno, M. (1992). *Raccoon II.* Tokyo: Chikuma Shobou.

Tajima, A. (1993). *Nyuugakusha senbatsuhouhou no kaizen* [Improving selection procedures for entering students]. In *Eigo kyouiku jittai chosa kennkyuukai* in JACET [Research on the current state of English education in JACET] (Eds.) *21 Seiki ni muketeno eigo kyouiku: Eigo kyouiku bessatsu, 1993* [English education towards the 21st century: English education supplement, 1993] (pp. 97-106). Tokyo: Taishuukan.

Wada, M., & Koike, I. (Eds.). (1990). *Kaitei koutou gakkou shidou youryo no tenkai, gaikokugo (eigo) ka-hen.* [Guidelines to revision in the high school curriculum, foreign language (English) course edition]. Tokyo: Meiji Tosho.

Writing Tools Group, Inc. (1992). *CorrectGrammar 3.0 for Macintosh.*

Zakaluk, B.L., & Samuels, S.J. (1988). *Readability: Its past, present, and future.* Newark, DE: International Reading Association.

Zenkoku tanki daigaku nyuushi mondai setkai, eigo•kokugo (92) [All Japan junior college entrance examination problem solutions, English-Japanese]. (1992). Tokyo: Obunsha.

(Received July 25, 1995; revised January 30, 1996)