

Readability in English Entrance Examinations

Laura MacGregor

Introduction

The purpose of this study is to evaluate the reading level of the reading passages on English entrance examinations at Gakushuin University over a five-year period (1999–2003 or Heisei 11–15). These reading levels will be compared with examinee performance on the accompanying marksheet items and also the difficulty of the translation items (English to Japanese) in these sections. Given that reading tends to dominate high school curricula, the fact that two reading passages are a consistent feature of each English examination (unlike most of the other sections of the tests, which tend to vary from year to year),¹⁾ as well as the fact that the reading passage items constitute half or more of the total exam points all confirm the importance placed on reading English texts. Considering the above, special care needs to be given in order to produce reading passages and test questions that best fit the level of the applicants.

The research questions for this study are as follows:

- (i) What are the difficulty levels of the reading passages of the Heisei 11–15 tests?
- (ii) How do they compare with the difficulty levels of the accompanying reading section items?
- (iii) How do they compare with the difficulty levels of the accompanying translation items?

The paper concludes with suggestions for future reading passages and

their accompanying test items.

Reading level is normally discussed as readability, defined simply as “the ease with which a document can be read” (Plain Language Center). Formulae to determine readability are mathematical equations that were first developed in the 1920s for use by textbook writers. According to Klare (1976, cited in Bruce and Rubin, 1988, p.6), since that time “hundreds have been proposed.” Readability scales are used as reading difficulty barometers in the preparation of government documents and insurance policies, and in many other domains.

The tools used in this study to evaluate reading level are two readability scales: the Flesch reading ease scale (developed in 1948 by Rudolf Flesch) and the Flesch-Kincaid grade level scale. They were chosen based on their availability and also the fact that they are most frequently used in language education research (Brown and Yamashita, 1995; Kimura and Visgatis, 1996; Alderson, 2000). Flesch reading ease (FRE) is calculated as follows:

$$\text{FRE} = 206.835 - (0.846 \times \text{NSYLL}) - (1.015 \times \text{W/S})$$

where NSYLL is the average number of syllables per 100 words and W/S is the average number of words per sentence (from Davies, 1984, p.188, cited in Alderson, 2000, p.71). This equation produces a number between 0 and 100 that expresses the difficulty level of the text on a scale in which the higher the number, the easier the text. Flesch recommends the 60–70 band as ideal, with a score of 65 being optimal (Flesch, 2003).

The Flesch-Kincaid grade level score, the other instrument used in this study, rates the text on a U.S. grade-school level. The formula is as follows:

$$\text{F-K} = (.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59$$

where ASL is the average sentence length and ASW is average number of syllables per word (Billingsley, 2004).

The Flesch reading ease scale (FRE) and the Flesch-Kincaid (FK) scale correlate as follows (Flesch, 2003) (Table 1):

Table 1: Correlation of Flesch Reading Ease and Flesch-Kincaid scales

<u>(FRE)</u>	<u>Prose Descriptor</u>	<u>FK U.S. School Grade</u>
90–100	very easy	5
80–90	easy	6
70–80	fairly easy	7
60–70	<i>plain English</i>	8-9
50–60	fairly difficult	10–12
20–50	difficult	college (30–50)
0–20	very difficult	college graduate (0–30)

Flesch uses the term “plain English” to describe his optimal range. This term has been widely applied by interest groups in several countries. A “Plain English” campaign launched in 1979 in Britain (Plain English Campaign), is a group of volunteers “fighting for public information to be written in plain English.” Other organizations also exist, among them, The Plain Language Action & Information Network, a U.S.-based government-wide group of volunteers who are working to improve communications from the federal government to the public (U.S. General Services Administration, 2003). Both groups emphasize the importance of being able to understand a text the first time it is read. They and others writing about plain English give tips for making writing clear and comprehensible. Among them are: (i) choose active verbs over passive verbs; and (ii) use short sentences rather than long ones.

Concerns about using readability scales

As a number of others have pointed out, relying exclusively on readability scales to judge the reading level of a text is overly simplistic

(Anderson and Davison, 1988; Bruce and Rubin, 1988; Kimura and Visgatis, 1996; Alderson, 2000). Some of the main criticisms are as follows:

First, since readability formulae measure only sentence length and word difficulty (based on word length), they cannot account for such elements as syntactic complexity, discourse cohesion, complexity of ideas, or reader characteristics which include motivation, interest, prior knowledge and the circumstances under which the passage is read (Bruce and Rubin, 1988, pp.7-8).

Second, Anderson and Davison (1988) cite several studies which show that sentence length does not necessarily affect understanding of a text, as Flesch (2003) and other advocates of plain English claim, and that length is not as crucial as complexity. For example, explicit conjunctions (i.e., because) make sentences longer, but facilitate comprehension for a wide range of readers of various ages and reading levels since they reduce sentence complexity (Anderson and Davison, 1988, p.35).

Third, on passage difficulty, Alderson (2000) states that, “Test designers should examine carefully the language of questions, rubrics and texts to ensure that they fall within the test population’s likely ability range” (p.81). He cautions against simplifying difficult tests because doing so will not only “disauthenticate the text, it also risks making the text harder to understand” (p.82). Instead, he suggests developing easier tasks or test questions to adjust for text difficulty (p.82). Conversely, “A reading score may be high or low because of item difficulty rather than text difficulty” (Alderson, 2000, p.86).

Finally, Alderson (2000, p.74) states that “readability formulae are rarely suitable for second- or foreign-language readers, even of English texts.” Clearly, this is an important consideration for this study and others like it and great caution should be taken when evaluating texts on readability formulae alone.

It is clear that, when constructing reading tests, careful attention needs to be paid to lexical, semantic, and syntactic aspects of the texts and their accompanying test items. However, despite the many concerns about the integrity of readability scales, the fact that they are widely used still remains. Therefore, they will be used here in this study, together with an examination of the reading test items to try to give a balanced and complete view of the difficulty of the reading passage tasks as a whole.

Materials and Method

The English entrance exams for five years (H11–H15) for all four faculties (law, economics, letters, and science) were examined. Since there are two reading passages on each test, a total of 40 reading passages and their accompanying test items comprise the data for this study. Microsoft 5.1a software for Macintosh (1992), which is equipped with the tools to calculate FRE and F-K and other textual features, was used for the mathematical calculations. Data from the entrance examination office at Gakushuin were obtained to examine individual item scores and overall test scores.

Results

The results will be presented in three sections covering three different aspects of the 40 reading passages according to the research questions: (i) the readability scales; (ii) the reading section items; and (iii) the translation items.

(i) Readability scales

The data for the tests were divided by faculty. Tables 2–5 present the readability data for the law, economics, letters, and science faculties respectively.

Table 2: Readability Data-Law

	H11-1	H11-2	H12-1	H12-2	H13-1	H13-2	H14-1	H14-2	H15-1	H15-2	average
Number of:											
words	328	257	275	336	425	312	474	380	594	524	391
characters	1933	1538	1666	2174	2532	1972	2640	2316	3494	3106	
paragraphs	3	3	4	4	4	5	6	3	7	6	
sentences	14	10	16	13	20	19	30	19	33	25	
Averages:											
sentences/paragraph	4	3	4	3	5	3	5	6	4	4	
words/sentence	23	25	17	25	21	16	15	20	18	20	
characters/word	4	4	4	8	4	5	4	4	4	4	
Readability:											
passive voice(%)	50	30	0	38	25	15	6	15	21	20	
Flesch reading ease	47.9	49.0	54.4	35.2	53.6	51.1	66.5	47.3	51.1	56.9	51.33
Flesch-Kincaid	12.3	12.7	9.6	13.9	11.0	9.6	7.6	10.7	10.1	9.8	10.42

Table 3: Readability Data-Economics

	H11-1	H11-2	H12-1	H12-2	H13-1	H13-2	H14-1	H14-2	H15-1	H15-2	average
Number of:											
words	260	371	392	455	397	491	518	302	626	382	419
characters	1595	2098	2135	2742	2410	2818	2992	1700	3378	2366	
paragraphs	4	3	4	5	6	7	5	2	10	7	
sentences	14	13	20	17	13	29	24	15	31	18	
Averages:											
sentences/paragraph	3	4	5	3	2	4	4	7	3	2	
words/sentence	18	28	19	26	30	16	21	20	20	21	
characters/word	4	4	4	4	4	4	4	4	4	4	
Readability:											
passive voice(%)	21	15	15	35	15	10	37	26	3	16	
Flesch reading ease	57.5	55.8	68.0	47.3	46.7	66.8	59.1	64.8	64.5	42.2	57.27
Flesch-Kincaid	9.0	12.5	8.6	12.6	12.5	7.7	10.3	9.2	9.2	12.6	10.42

Table 4: Readability Data-Letters

	H11-1	H11-2	H12-1	H12-2	H13-1	H13-2	H14-1	H14-2	H15-1	H15-2	average
Number of:											
words	324	306	363	390	295	491	481	415	338	459	386
characters	2027	1799	2088	2356	1844	2014	2796	2239	1877	2736	
paragraphs	2	4	4	4	4	5	4	4	5	5	
sentences	15	14	16	16	16	25	24	21	20	28	
Averages:											
sentences/paragraph	7	3	4	4	4	5	6	5	4	5	
words/sentence	21	21	22	24	18	19	20	19	16	16	
characters/word	5	4	4	4	5	4	4	4	4	4	
Readability:											
passive voice(%)	26	21	12	12	25	11	20	0	10	10	
Flesch reading ease	46.5	51.3	57.4	44.7	43.5	43.8	54.6	75.9	65.1	59.3	54.21
Flesch-Kincaid	10.7	11.5	10.2	12.7	11.7	12.0	10.4	6.9	8.1	8.9	10.31

Table 5: Readability Data-Science

	H11-1	H11-2	H12-1	H12-2	H13-1	H13-2	H14-1	H14-2	H15-1	H15-2	average
Number of:											
words	376	324	419	359	302	343	428	305	395	309	356
characters	2937	1945	2416	2056	1856	2104	2609	1897	2304	1856	
paragraphs	3	1	4	7	3	7	3	2	4	6	
sentences	17	15	18	27	17	15	19	12	19	17	
Averages:											
sentences/paragraph	5	15	4	3	5	2	6	6	4	2	
words/sentence	22	21	23	13	17	22	22	25	20	18	
characters/word	4	4	4	4	4	4	4	5	4	4	
Readability:											
passive voice(%)	17	26	5	18	35	13	10	41	10	5	
Flesch reading ease	64.3	49.9	56.6	75.2	48.6	45.4	44.8	35.6	59.6	61.3	54.13
Flesch-Kincaid	9.5	9.5	9.5	9.5	10.6	12.2	11.8	14.5	9.3	8.9	10.55

Looking at the entire data corpus of 40 passages, the average passage length is 391 words with texts ranging in length from 257 words to 626 words. That these two passages have FREs of 49.0 and 64.5 respectively indicates that passage length does not contribute to reading ease.

The average FRE is 54.24 with a range of 35.2–75.9, and the average F-K is 10.5 with a range of 6.0–14.5 (both low and high scoring texts are found in Table 5, the science faculty). The letters and science faculty passage average FREs are closest to the global average of 54.24, with a small difference in the FRE for law (51.33) and for economics (57.27). However, all of the averages are within the 50–60 band, meaning that they are all within the “fairly difficult” range according to Flesch’s scale of high school 10th–12th grade levels.

There are 11 passages in which the percentage of passive voice content is 10% or less. Of these, the FRE is higher than the average of 54.24 for all texts except one, (Table 5, H14-1 with an FRE of 44.8.). This suggests that low frequency of passive voice results in easier texts. However, in the two instances where there is no passive voice used at all (Table 2, H12-1 and Table 4, H14-2) the FREs are 54.4 and 75.9. Since they are radically different (a 21.5 point spread) and one is below Flesch’s optimal range, it cannot be concluded, at least from these data, that passive voice necessarily makes a text easy or difficult.

Sentence length as a measure of passage difficulty, another point of contention among advocates of plain English, was also examined. Of the 15 passages containing sentences with an average of less than 20 words, seven of them had FREs less than 60. Thus, it cannot be concluded from these data that shorter sentences are easier than longer sentences. Other factors, such as vocabulary, syntax, and cohesion are obviously at work in determining sentence difficulty.

(ii) Reading section items

The number and type of items in the reading sections and the weighting they are allotted in terms of the entire test are illustrated in Table 6. The first column shows the weighting of the reading sections on the tests (the two reading passages combined). For example, the reading sections carried 90 points for the H11 law exam, constituting 60% of the entire test. The number of points for the reading sections comprise half or more of the total points of the test, or 49.3%–70.0%.

The second column shows the total number of test questions for the two reading passages combined. There are between 10–20 items in total in the two reading sections on each exam. They can be divided into two types, receptive and productive, as shown in the third column.

Table 6: Reading Section Items

	weighting points(%)	no.of items	receptive/ productive	productive items no.of points(%)
Law				
H11	90(60.0%)	10	5(5)	39(26.0%)
H12	85(56.6%)	14	10(4)	27(18.0%)
H13	95(63.3%)	19	12(7)	44(29.3%)
H14	85(56.6%)	15	10(5)	35(23.3%)
H15	77(51.3%)	18	15(3)	23(15.3%)
Economics				
H11	78(52.0%)	15	9(6)	45(30.0%)
H12	90(60.0%)	17	12(5)	39(26.0%)
H13	77(51.3%)	19	13(6)	28(18.6%)
H14	80(53.3%)	21	17(4)	31(20.6%)
H15	83(55.3%)	20	16(4)	28(18.6%)

	weighting points(%)	no.of items	receptive/ productive	productive items no.of points(%)
Letters				
H11	90(60.0%)	11	5(6)	64(42.6%)
H12	91(60.6%)	18	12(6)	40(26.6%)
H13	74(49.3%)	14	9(5)	29(19.3%)
H14	77(51.3%)	14	10(4)	30(20.0%)
H15	86(57.3%)	19	17(2)	24(16.0%)
Science				
H11	70.0%	14	9(5)	35.00%
H12	65.0%	14	8(6)	36.00%
H13	55.0%	13	8(5)	34.00%
H14	55.0%	14	10(4)	23.00%
H15	64.0%	17	13(4)	23.00%

note: Law, Economics, and Letters tests are out of 150 points Science test is out of 100 points

The receptive items, which are multiple-choice items on these tests, consist of the following item types: single-word synonyms (in English or in Japanese), phrasal synonyms, reading comprehension (in English or in Japanese), and single-word accent or pronunciation identification (discontinued from H15). The productive items are of the following item types: full or partial sentence or short phrase translation from the reading text into Japanese, writing a word or phrase in English to complete a task (usually comprehension-based, such as “What two words in the passage do the underlined words refer to?”), and explaining or paraphrasing a phrase from the text in Japanese. Apart from two exceptions (law H13-2 and economics H14-2), there was at least one English-to-Japanese translation or explanation/paraphrasing item for each reading passage on each test, showing that testing translating or interpreting from English to

Japanese is routine. The last column shows the number of points in the reading section that were allotted for the productive items. Again, looking at the H11 law test, 39 points (out of 90) were for productive items, 26% of the test.

Data were not available for the productive items on the tests, and data on marksheet items were available beginning in H14, since the university only began to collect this type of data (unpublished) from 2002.

For the H14 and H15 exams, item facility data furnished by the university entrance exam section (unpublished) can be presented and discussed here. Item facility refers to the percentage of examinees who answer an item correctly. It is expressed as a value of 1. Items for tests such as entrance exams should ideally have an average IF of .50 (indicating that 50% of the test population got the item correct), and items in the range of .30–.70 are usually considered acceptable (Brown, 1996, pp.69–70). There were a number of reading section marksheet items on each of the tests in H14 and H15 that fell outside of the acceptable range. They averaged five items per test and most of them were on the easy end of the scale (above .70).

In order to observe the relationship between item facility and reading ease, two examples will be taken, one of a very difficult text and one of a very easy text from this sample. The first is a reading passage from a science exam (H14–2). There were six items for this reading passage, five of which were receptive and one which was productive. It is the most difficult reading passage in the entire sample, with an FRE of 35.6 and an F-K level of 14.5, indicating that this text would be best suited for a 2nd-year university student in the U.S. Of the six items, four of them can, for the most part, be answered without consideration of the content or overall comprehension of the text. They are as follows: (i) choose the

definition for the underlined word from the text (IF=83.2); (ii) choose the meaning of the phrase (in English) that does not match the phrase in the text (IF=47.7); (iii) choose the phrase (in Japanese) that best matches the meaning of the phrase in the text (IF=55.5); and (iv) translate the phrase in the text into Japanese (IF not calculated). The two text-dependent items ask the following: (v) to answer a comprehension question about the text (in Japanese; IF=19.8); and (vi) to choose the best title for the passage (in English; IF=63.2). The comprehension item (v) was the only difficult item in this section (IF=19.8) and the definition item (i) was very easy for examinees (IF=83.2). The other four items were within the acceptable IF range.

The second example uses the easiest reading passage in this sample, the second text in a letters exam (H14-3). Again, there were six items, four receptive and two productive. The FRE was 75.9 and the F-K level was 6.9, making the passage suitable for a student nearly ready to begin junior high school (grade seven). Five of the items could be answered independently of the text: (i) two sentence gap-fills (IF=74.3; 55.8); (ii) a “choose the correct phrasal synonym” item (in English; IF=42.7); (iii) one which involves choosing the phrase (in Japanese) which best matches the meaning of the phrase in the text (IF=18.2); (iv) a phrase stress identification item (IF=28.8); and (v) an English to Japanese translation (IF not calculated). The text dependent item is: “explain the underlined section in the text in Japanese” (IF not calculated). The first part of the first item was a very easy item, and there were two difficult items, both text independent (iii and iv above).

(iii) Translation items

Of the 40 reading passages, 33 translation items could be analyzed (seven had to be discarded because they were short phrases rather than

full sentences or near full sentences). Even of the remaining 33, the statistics may be invalid, since the unit of analysis was a single sentence averaging less than 20 words. Given that readability instruments are recommended for texts of 200 words or more, the following data should be interpreted cautiously.

The average number of words in the sentences for translation was 17.1; the average FRE was 68.1 (with a range of 39.5–95.7), and the average F-K was 7.8 (with a range of 2.4–15.6). Recalling that the average FRE and F-K for the complete passages are 54.24 and 10.5, the reading levels for these sentences appear to be much easier than those of the complete texts, since their difficulty is nearly three grade levels lower.

Discussion

(i) Readability

Brown and Yamashita (1995) examined the reading passages on the 1993 English exams of 10 top ranking private universities²⁾ in Japan. Kimura and Visgatis (1996) did a study of 33 junior college entrance exams from 1992 and compared them with 66 texts from four senior high school English II readers used in the 1992 academic year (this text level was chosen because it fit the prerequisite profiles for the junior college tests they examined). Their data are shown with the data for Gakushuin collected here in Table 7. It should be noted that there is a 10-year span between the Gakushuin data and the other data and that Gakushuin's tests of 10 years ago may have been different from those being reported here.

Table 7: Readability in three studies

	<u>10 PUs</u>	<u>Gakushuin</u>	<u>33 JC</u>	<u>4 SHS</u>
	<u>(1993)</u>	<u>(1999–2003)</u>	<u>(1992)</u>	<u>(1991–1992)</u>
av. no of words (range)	540 (381–986)	388 (257–626)	N/A	N/A
FRE (range)	60.4 (48.1–65.5)	54.24 (35.2–75.9)	64.8 (41.9–92.9)	75.99 (46.0–98.3)
F-K (range)	9.4(8.2–11.3)	10.5(6.9–14.5)	8.3(1.6–13.2)	5.96(1.3–11.8)

Notes: Kimura and Visgatis did not report the word counts in their study. For the sake of consistency, all data were rounded off to one integer.

First, the Gakushuin passages are generally much shorter than those of other private universities (PUs). Second, they have the lowest average FRE, with a much wider range than that of the 10 private universities (a 40.7 spread versus the 17.1 spread of the 10 university texts), but a smaller range than the junior college (JC) and senior high school (SHS) texts (by about 10 points). Third, the F-K range for the Gakushuin texts is wider than that of the 10 universities, which have the narrowest range (only three grade levels, compared with eight grade levels for Gakushuin and more than ten grade levels for the junior college and senior high texts). The data for the 10 universities are the tightest, most consistent of the four groups. That they have achieved this commonality shows that somehow they are in agreement about the level at which applicants to their universities should be reading on the tests.

Generally, there are slightly different expectations from faculty to faculty of students' English language ability. These expectations are reflected in the varying demands of the English classes that the faculties require their students to take. Therefore, differences in the level of difficulty on the reading passages and their accompanying test questions

should be expected between faculties. However, wide spreads in difficulty level from year to year within each faculty presently exist and they need to be addressed. Taking the law faculty exams as an example, over the five years that this study covers, the FREs range from 35.2 to 66.5, which translates into a spread of seven U.S. grades (7.6–13.9). If a target level of reading difficulty could be established for each faculty, it would help both test makers and test takers prepare for the tests and achieve better results.

Another point concerning difficulty level is that there is no difference in the incoming requirements for applicants, since the university states the following prerequisites for the English exams across all four faculties: English I, English II, Reading, and Writing (Gakushuin Daigaku, 2003). Thus, students are not likely to know the target level of difficulty of the test unless they look specifically at past exam papers, and even then, they will be faced with a wide range of difficulty levels from year to year. For the above reasons, a target level of difficulty needs to be established for each faculty in order to achieve consistency on the exams.

The level of the high school reading texts are much easier than any of the entrance exams shown in Table 7. However, to target student reading level on the entrance exam based solely on textbook contents would be a mistake, since students are exposed to other reading materials in high school, particularly during the months preceding the entrance exams, which are likely to be more difficult. They may attend intensive exam preparation lessons at school once their formal instruction ends in December of their third year of high school, and may also attend cram school, which use entrance exam difficulty levels or higher as target levels of instruction. On the other hand, if the reading levels or test item levels are too high on entrance exams, examinees may guess and by doing so weaken the reliability of the test (Kimura and Visgatis, 1996, p.90). Establishing and maintaining a target difficulty level is necessary.

Further complicating the selection of the most appropriate target level is the fact that there is no standardized vocabulary list of what high school students are sure to have covered (as there is for junior high school). Even dictionaries do not agree on the vocabulary for high school or above high school level. Thus, it is important for university exam item writers to consider as many factors as possible when determining difficulty levels and preparing the tests.

Finally, the topic of a reading passage may affect its readability, since knowledge of a certain topic and its associated vocabulary may affect reading ease, overall comprehension, and performance on the questions. In this sample, it is not known how or whether a topic affects readability. Looking at the passage with the lowest FRE (law H12-2), the topic is the language of the law, certainly relevant to prospective law majors, but may not be a familiar subject for them as high school students. Also at work here is the style of the passage and the level of the vocabulary: this is a sophisticated academic text that may require more background knowledge than test takers are equipped with. The following sentence from the text is an example of this point, and is especially significant because a test question is based on the underlined part (italics in original):

“The repetition and rhythm of many expressions (for example, *the truth, the whole truth...*) reflect the need, in an age before printing and general literacy, for the law to be remembered clearly, and passed on consistently.”

In contrast is the passage with the highest FRE (letters H14-2), which is a narrative of a familiar scene: an impatient mother and her small son who is struggling with the task of tying his shoelaces. While topical knowledge may have some effect on the readability of the passage, the extent of its influence is beyond the scope of this study.

(ii) Reading section items

According to Table 7, generally speaking, the more test questions there are for a reading passage, the fewer productive items there are among them. Where there are relatively few questions, the receptive-productive ratio is often nearly 1:1. Certainly, this is fair, since productive questions (particularly paraphrasing and translating tasks) require more time than do multiple choice questions. The point is, however, what kinds of questions, receptive or productive, better assess the language abilities of the applicants? Why are there so many text-independent questions in the reading sections? Of the 12 items in the two reading passages presented in the results section above, nine of them are language questions that can be answered independently of the text. If reading skills are to be evaluated fully, more items that are relevant to the text need to be included.

Looking at the far right column of Table 6, it would appear that productive items in the reading passages alone figure anywhere from 15% to 36% of the total exam score. In addition, half of the tests have productive item type sections testing grammar (section III in each of the following: law H11 and H15, economics H12–15, letters H11, H13, H15, and science H13), and all tests have a two-sentence Japanese to English translation question at the end, which carries 6.7% for the law, economics, and letters exams, and 10% for the science exam. Thus, the presence of productive items on the tests in their entirety for the whole sample averages 35.45%. When compared with the 11.37% for the 10 private universities' tests in Brown and Yamashita's study (1995, p.25), it would appear on the one hand that the Gakushuin exams pay much more attention to productive item types than the norm. However, nearly all of the productive items on the Gakushuin tests are translation tasks (either Japanese to English or English to Japanese), and given that "translation

was abandoned years ago in ESL instruction” (Brown and Yamashita, 1995, p.28), it appears that Gakushuin is maintaining an archaic tradition that needs to be reexamined. A chief reason why translation has lost favour among Western second and foreign language educators is simple: translation is a highly specialized skill for which university applicants who have officially had only six years of formal English education, albeit with some translation practice during that time, are not equipped to handle (Brown and Yamashita, 1995, p.11). Therefore, it is neither practical nor realistic to assign translation tasks on such a high stakes test as the entrance exam.

(iii) Translation items

Concerns regarding the validity of translation tasks have already been outlined in the previous section. Since the translation items in the last section of the exam are for Japanese to English, it is not possible to compare their level of readability with those in the reading passages. Nor is it possible to determine the correlation between reading ease and translation ease. Therefore, while the sentences for translation in the reading passages appear to be relatively easy to read (in comparison with the rest of the texts in their entirety), whether they are easy to translate or not may depend on other factors. Given the complications involved in translation, the reevaluation of its presence in both the reading section and the final section of the tests should be considered.

Conclusion

This paper examined 40 English reading passages and their accompanying test questions on the English entrance exams at Gakushuin University for the years 1999–2003. The readability of the texts was assessed using the Flesch Reading Ease and Flesch-Kincaid scales. Their

average difficulty level corresponded to U.S. grade 10 (first-year senior high school) reading level for native speakers of English. This level was somewhat higher than similar studies of university and junior college entrance exam reading passages, and significantly higher than the reading passages in senior high school English II textbooks. In closing, a number of points for consideration will be presented with the view of improving readability and overall test quality in the future.

The global average FRE was 52.24 and average FREs for each faculty fell within the 50–60 “fairly difficult” band. If this is a desirable difficulty level, effort should be made in the future to select reading passages which reflect this difficulty band more closely. In any case, a standard target readability range for each faculty’s exam should be established and agreed upon in order to produce consistent tests from year to year and to make it clear to applicants what is expected of them on the test. A standard for text lengths should be established as well.

When selecting reading passages, other factors which contribute to readability should be taken into account. Among them are: (i) vocabulary level, (ii) semantic complexity, (iii) syntactic complexity, and (iv) discourse cohesion.

Concerning the test items which accompany the reading texts, consideration should be given in the future to the balance between receptive and productive items and the number of text-dependent items versus the number of text-independent items to include. Before that, the issue of whether or not to include items which can be answered independently of the text should be addressed. In the end, item difficulty should complement the reading difficulty of the texts in order to produce a balanced picture of applicants’ reading and comprehension skills. The inclusion of translation tasks in the reading section and elsewhere on the exam merits review.

Future research should examine the role that topic and type of discourse play in readability. This would assist test makers in selecting appropriate reading texts for testing purposes. All of the above suggestions are being introduced as a result of this preliminary readability study. Clearly, more work needs to be done in this area to gain an even clearer picture of what readability is and how it can be best assessed with learners of English as a foreign language.

Notes

- 1) The only other constant section on the tests is the final item, a two-sentence from Japanese to English translation task, worth 10 points on the exams of all four faculties.
- 2) The ten universities are: Aoyama Gakuin, Doshisha, Keio, Kansai Gaidai, Kansai, Kyoto University of Foreign Studies, Rikkyo, Sophia, Tsuda, and Waseda (Brown & Yamashita, 1995, p.12).

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Anderson, R.C., & Davison, A. (1988). Conceptual and empirical bases of readability formulas. In A. Davison and G.M. Green (eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered* (pp.23-54). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Billingsley, D.B. (2004, March 3). Flesch-Kincaid reading measures. [Online]. Available: <http://www.louisville.edu/a-s/english/dale/601/flesch.htm> [2003, August 16].

Brown, J.D., & Yamashita, S.O. (1995). English language entrance examinations at Japanese universities: What do we know about them? *JALT Journal*, 17 (1), 7–30.

Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.

Bruce, B. & Rubin, A. (1988). Readability formulas: Matching tool and task. In A. Davison and G.M. Green (eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered* (pp.5–22). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Davies, A. (1984). Simple, simplified and simplification: What is authentic? In J. C. Alderson and A.H. Urquhart (eds.), *Reading in a foreign language*. London: Longman.

Flesch, R. (2003). How to write plain English. University of Canterbury, NZ, Department of Management page link. [Online]. Available: <http://www.mang.canterbury.ac.nz/courseinfo/AcademicWriting/Flesch.htm> [2003, August 16].

Gakushuin Daigaku. (2003). [Online]. Nyuushi jouho (Entrance exam information). Available: <http://www.gakushuin.ac.jp/univ/adm/adm/> [2003, August 17].

Kimura, S., & Visgatis, B. (1995). High school English textbooks and college entrance examinations: A comparison of reading passage difficulty. *JALT Journal*, 18 (1), 81–95.

Klare, G. R. (1976). A second look at the validity of readability formulas. *Journal of Reading Behavior*, 8, 129-152.

Microsoft Corporation. (1992). *Microsoft 5.1a for Macintosh*.

Plain English Campaign. (date unknown). Introduction to plain English campaign. [Online]. Available: <http://www.plainenglish.co.uk/introduction.html> [2003, August 16].

Plain Language Center. (date unknown). Everything you ever wanted know about readability tests but were afraid to ask. [Online]. Available: <http://www.gopdg.com/plainlanguage/readability.html> [2003, August 16].

U.S. General Services Administration. (2003, April 14). Plain language. [Online]. Available: <http://www.plainlanguage.gov/> [2003, August 16].

英語入試問題におけるリーダビリティについて

ローラ マクレガー

本研究の目的は、平成 11-15 年（1999-2003）の 5 年間にわたる学習院大学入学試験（英語）の内容読解問題の問題文のレベルを検討することであり、その際に比較される対象は、マークシート問題と英文和訳問題の難易度である。

本学では、英文読解が高校教育課程の根幹であるとの前提のもと、他の出題形式は年度により変更があるのに対し、2 題の内容読解問題は毎年必ず出題されている。このことは、内容読解問題の得点が総得点の半分以上を占めるという事実と合わせて見た場合、本学の入学試験において、英文の「読解」がいかに重視されているかを裏付けるものである。故に、内容読解問題の作成において、受験者の水準に適切な英文を用意することには、細心の注意を払う必要がある。

本研究における調査項目は以下の通りである。

- (i)平成 11-15 年の内容読解問題の問題文の難易度レベルはどうであるか。
- (ii)その難易度レベルは、マークシート項目の難易度レベルと比較してどうであるか。
- (iii)その難易度レベルは、英文和訳問題の難易度レベルと比較してどうであるか。

さらに、論文の末尾に、今後の内容読解問題用英文の難易度レベルと設問項目についての私見を付す。