

The Miyazaki EFL Readability Index

Jerry Greenfield

フレッシュ・キンケイド難易度推定式はMICにおいてEFLの授業で使用する教材の読解難易度の判定や、その他の目的のためによく使われる。フレッシュ・キンケイド難易度推定式やその他の、英語を母国語とする人達のために作られたものが、MICの学生に適しているかどうかを判定する調査を、1998年に200名のMICの学生を対象にして行った。それらの推定式は、我々の学生に対しても有効であることが判明した。また今回の調査で、特にMICの学生に有効な推定式を作り上げることも可能なことがわかった。この新しい推定式・宮崎EFL読解難易度推定式と呼ばれるものはフレッシュ難易度推定式によく似ており、100点満点で単語と文章の長さを計測する。簡単な解釈表があり、宮崎EFL読解難易度推定式はとても使いやすい。これらの研究に追隨して、実際にMICで使用される教材の推定式の結果が学生の読解力にどういう意味があるのかを問う研究が、開始された。

The Flesch and Flesch-Kincaid readability scales are regularly used at MIC to gauge the reading difficulty of EFL materials for use in our classrooms and other purposes. To verify whether these and other classic measures developed for native-English speakers are valid to use in our situation, a study was conducted in 1998 with 200 MIC students. The results conclusively showed that the formulas are valid to use with our students. It also was possible to produce a new formula specifically calibrated for MIC's students. This formula, called the Miyazaki EFL Readability Index (MEFLRI), uses word and sentence length to calculate scores on a 100-point scale similar to the Flesch Reading Ease. A simple look-up table makes the formula easy to use. Additional exploratory research has begun to shed some light on what these scores mean in terms of student reading performance with texts as they are actually used at MIC.

Readability Formulas: Tools are there, but are they right for EFL?

Content-based teaching requires teachers to select and create appropriate reading materials which their students can successfully read. Teachers at MIC have the added challenge that our teaching materials need to be in English that can be read by native speakers of Japanese. With few content-area textbooks written especially for EFL use, and perhaps none that do exactly what we need, we must sift through mounds of texts looking for ones we think we can use. In the end, we are likely either to adapt what is there or surrender to writing our own materials from scratch. In any case, it is hard to predict how difficult our students will find what we give them to read. Looking for help, many of us have taken advantage of the readability tools available in Microsoft Word. With some guesswork, perhaps advice from colleagues, and our own experience, we settle on a range of readability scores as a workable target for texts that are appropriate for the various tasks we expect our students to perform with them. Thus, after joining the MIC faculty in 1997 I quickly learned where to find the readability report in my word processor.

Readability scales have been used for these purposes since the 1930s. They are based on multiple-regression formulas derived by testing a group's reading comprehension of a series of texts and relating the comprehension scores to some measurable variables in the texts. The factors most often used are lexical difficulty and grammatical difficulty, typically measured as word length and sentence length. These in fact are the two variables used in the Flesch Reading Ease (Flesch, 1948) and Flesch-Kincaid Grade Level (Kincaid et al., 1975), the most popular of the many readability scales, and the two which are included in current versions of Microsoft Word. Readability scales have been criticized because they do not seem to take due

account of the psycholinguistic complexities of the reading process (see, for example, Bruce and Rubin, 1988.) Yet the formulas continue to be widely used, because, face validity aside, they have been shown time and again to be remarkably accurate in predicting the difficulty readers will have in reading texts (for an overview of validation studies, see Chall and Dale, 1995).

However, as I learned more about these formulas and how they were developed, I wondered whether it was really appropriate to use them at MIC with our Japanese-speaking students. After all, the classic formulas were constructed from test results with American native-English speakers, usually school students. Furthermore, the many studies that have revalidated the formulas were also conducted with the same kind of readers. Considering differences in how second languages are learned—not to mention differences in cultural background and age differences between native-speakers and second-language speakers at comparable English levels—should I assume that the readability formulas are reliable in our EFL setting?

A college-wide reading assessment at MIC in 1998 offered an opportunity to dispel my doubts on this issue. Reading test scores from 200 first- through fourth-year MIC students over a set of thirty-one standard passages were correlated with readability scores on the Flesch, Flesch-Kincaid, and several other classic readability measures. The results showed clearly that the formulas do in fact predict the relative difficulty of English passages as accurately for our students as they do for native speakers (Greenfield, 1999). These results were especially convincing because the reading test used in the assessment was based on one used in a standard study by Bormuth (1971), which has been extensively cross-validated with the classic formulas many of us were using everyday at MIC. This is important, because it means we can apply these formulas confident that the information they give us about reading difficulty is basically as valid for our students as it is for native-English speakers.

A readability index for EFL

But what exactly is the information the formulas are giving us? Statistically validating the formulas is one thing; knowing what their scores mean in practice is actually quite another. The Flesch Reading Ease formula gives us a score on a 100-point scale, with 50 indicating average difficulty for native-speakers of English, 100 being extremely easy, and zero, presumably, impossibly difficult for anyone. A conversion formula is available to relate the Reading Ease score to an American school grade level, although Microsoft Word supplies only the scaled Reading Ease score. The Flesch-Kincaid formula, on the other hand, yields a grade-level directly. When the formula reports that a particular passage has a tenth-grade readability, it is predicting that a typical tenth-grade American class would score an average of 75% on a well-constructed reading comprehension test over that passage. That may be useful information for American school teachers, but our students are not American school students, so how can we relate that score to what our students could do with that passage? What is an acceptable level of performance for our students? Is 75% accuracy on a multiple-choice test a realistic and appropriate criterion for judging reading performance by EFL students? Does the way we are using the reading matter? Does the amount of support we provide matter? We can

Figure 1. Miyazaki EFL Readability Index Look-up Table

Words per Sentence	Letters per Word																									
	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0	5.1	5.2	5.3	5.4	5.5	5.6	5.7
2	99	97	95	93	92	90	88	86	84	82	80	78	77	75	73	71	69	67	65	63	62	60	58	56	54	52
3	99	97	95	93	92	90	88	86	84	82	80	78	77	75	73	71	69	67	65	63	61	60	58	56	54	52
4	97	95	93	91	90	88	86	84	82	80	78	76	75	73	71	69	67	65	63	61	60	58	56	54	52	50
5	95	93	91	90	88	86	84	82	80	78	76	75	73	71	69	67	65	63	61	60	58	56	54	52	50	48
6	93	91	90	88	86	84	82	80	78	76	75	73	71	69	67	65	63	61	59	58	56	54	52	50	48	46
7	91	90	88	86	84	82	80	78	76	74	73	71	69	67	65	63	61	59	58	56	54	52	50	48	46	44
8	89	88	86	84	82	80	78	76	74	73	71	69	67	65	63	61	59	58	56	54	52	50	48	46	44	42
9	88	86	84	82	80	78	76	74	73	71	69	67	65	63	61	59	57	56	54	52	50	48	46	44	42	41
10	86	84	82	80	78	76	74	72	71	69	67	65	63	61	59	57	56	54	52	50	48	46	44	42	41	39
11	84	82	80	78	76	74	72	71	69	67	65	63	61	59	57	56	54	52	50	48	46	44	42	41	39	37
12	82	80	78	76	74	72	71	69	67	65	63	61	59	57	56	54	52	50	48	46	44	42	40	39	37	35
13	80	78	76	74	72	70	69	67	65	63	61	59	57	55	54	52	50	48	46	44	42	40	39	37	35	33
14	78	76	74	72	70	69	67	65	63	61	59	57	55	54	52	50	48	46	44	42	40	39	37	35	33	31
15	76	74	72	70	69	67	65	63	61	59	57	55	54	52	50	48	46	44	42	40	38	37	35	33	31	29
16	74	72	70	69	67	65	63	61	59	57	55	53	52	50	48	46	44	42	40	38	37	35	33	31	29	27
17	72	70	68	67	65	63	61	59	57	55	53	52	50	48	46	44	42	40	38	37	35	33	31	29	27	25
18	70	68	67	65	63	61	59	57	55	53	52	50	48	46	44	42	40	38	36	35	33	31	29	27	25	23
19	68	67	65	63	61	59	57	55	53	51	50	48	46	44	42	40	38	36	35	33	31	29	27	25	23	21
20	66	65	63	61	59	57	55	53	51	50	48	46	44	42	40	38	36	35	33	31	29	27	25	23	21	20
21	65	63	61	59	57	55	53	51	50	48	46	44	42	40	38	36	34	33	31	29	27	25	23	21	19	18
22	63	61	59	57	55	53	51	49	48	46	44	42	40	38	36	34	33	31	29	27	25	23	21	19	18	16
23	61	59	57	55	53	51	49	48	46	44	42	40	38	36	34	33	31	29	27	25	23	21	19	18	16	14
24	59	57	55	53	51	49	48	46	44	42	40	38	36	34	33	31	29	27	25	23	21	19	17	16	14	12
25	57	55	53	51	49	48	46	44	42	40	38	36	34	32	31	29	27	25	23	21	19	17	16	14	12	10
26	55	53	51	49	47	46	44	42	40	38	36	34	32	31	29	27	25	23	21	19	17	16	14	12	10	10
27	53	51	49	47	46	44	42	40	38	36	34	32	31	29	27	25	23	21	19	17	15	14	12	10	10	10
28	51	49	47	46	44	42	40	38	36	34	32	30	29	27	25	23	21	19	17	15	14	12	10	10	10	10
29	49	47	45	44	42	40	38	36	34	32	30	29	27	25	23	21	19	17	15	14	12	10	10	10	10	10
30	47	45	44	42	40	38	36	34	32	30	29	27	25	23	21	19	17	15	13	12	10	10	10	10	10	10

Note: 50 = average difficulty (gray boundary). Above the diagonal is easier, below more difficult.

think of more such questions, but they all underscore the point that readability scores are good only for distinguishing the relative difficulties of different texts. Knowing whether our students can do what we want them to do with a text at a given readability level must depend on our purposes and experience using texts at that level. We now know that the formulas do let us compare the difficulties of texts in a way that is objective and reliable. Still, to be really useful the readability scores require experienced interpretation.

Fortunately, using the data used from the 1998 reading assessment it was possible to produce a new formula, called the Miyazaki EFL Readability Index (MEFLRI), using the same text variables as the Flesch and Flesch-Kincaid formulas but adjusted to reflect the reading performance of our students. The new formula is as follows:

Miyazaki EFL Readability Index (scaled)

$$\text{EFL Difficulty} = 164.935 - (18.792 \times \text{Letters per Word}) \\ - (1.916 \times \text{Words per Sentence})$$

The new index reports scores on a 100-point scale similar to the Flesch Reading Ease, indicating the relative difficulty of texts without reference to performance criteria such as those implied in grade leveling. A score of 50 on the Miyazaki Index represents a text of average difficulty for MIC students.¹ I have argued that the sample of students who provided data for this analysis was typical of Japanese EFL students in general so that the index's validity can be generalized to Japanese EFL readers, and probably also to EFL students whose native language is other than Japanese (Greenfield, 1999). While the scores may not mean exactly the same across language backgrounds, nevertheless it seems likely that the scale is probably more appropriate to EFL students in general than is the Flesch scale and certainly more appropriate than the grade levels of the Flesch-Kincaid and other formulas.

A look-up table (Figure 1) plotting word length against sentence length makes the Miyazaki Index simple to use. These counts are easy to get for digital texts in most word-processing applications or to count by hand for printed texts. Texts totaling less than 300 words and digital texts of any length should be counted in their entirety. To hand-count word and sentence lengths of longer texts, it is customary to select and count three 100-word samples from near the beginning, middle, and end of the text. Sentence lengths are computed by dividing the word count by the number of sentences and rounding to the nearest whole number. Word length is calculated by dividing the total number of characters by the number of words and rounding to the nearest tenth. The word length and sentence length values can then be located in the look-up table and the score at that row and column intersection is the readability score for the text. Notice that the score of 50 forms a diagonal from the bottom left to upper right. Higher scores to the left of the diagonal represent below-average difficulty, lower scores to the right above average.

¹ This score was associated with a fifth-word deletion cloze score of about 26%.

Relating scores to performance criteria: What do the scores really mean?

Scaling readability scores to reflect the characteristics of our EFL students eliminates one interpretive step, but I still wondered what the scores mean in terms of practical performance. Which students can successfully read a text with a readability of 50? Indeed, what should we understand *is* successful reading of such a text? If the conventional criterion is a score of 75% on a comprehension test, is that an appropriate criterion for successfully reading EFL texts that we assign students in, say, first-year classes at MIC? If not, then what criterion should we use, and how would it relate to readability scores, whether measured by the new Miyazaki Index or by the classic formulas?

A partial answer to this question is provided by supplemental data from the 1998 reading assessment. After reading each of the 31 test passages, students were asked to rate on a three-step scale the difficulty of the passage for use in an MIC class. The middle rating of "about right" was associated with an MEFLRI score in the mid-50s—a little easier than the mid-point on the index. This finding was confirmed by data from a subsequent college-wide reading assessment in 2000, in which 257 MIC students read three passages spanning the middle difficulty range found in the 1998 assessment. Again the students rated the difficulty of each passage, this time on a five-point scale, on which the middle rating of "about right" was associated with an MEFLRI score of 57, slightly easier than but generally consistent with the rating indicated in the first study. From a strict research standpoint, the students' ratings are subjective reports that could indicate a mix of undefined and uncontrolled-for factors ranging from simple comfort level to pre-formed expectations about college texts. Nonetheless, we may conclude from this composite indication of students' perceptions, that an MEFLRI score of 50 for a text indicates that our students on average will find that text "somewhat hard," but not too hard.

I looked at the meaning of readability scores from another angle in a small study in 2000 in which I applied the Flesch, Flesch-Kincaid, and Miyazaki measures to representative academic texts used in MIC courses (Greenfield, 2001). For this purpose, a total of forty passages were submitted by eight members of the MIC faculty representing the fields of history, religion, art history, philosophy, psychology, education, and economics. Nineteen of the passages were used in courses taught in the first year or first semester of the second year, and twenty-one were used in upper division courses for combined third- and fourth-year students. (All MIC students are engaged in study abroad in English-speaking countries during the second semester of the second year. Thus the lower division spans three instead of four semesters.)

Text contributors were asked to supply passages which in their view had been successfully read in a normal way by their students. "Normal" in this context meant that students were expected to read and study the text essentially on their own, typically with some pre- and/or post-reading support in class and/or collateral apparatus such as a glossary or study guide, but otherwise without close support by the instructor. What I had in mind here is basically the middle condition in the following traditional three-tier classification of conditions for reading (Tinker 1965, Bormuth 1968):

Reading condition	Cloze accuracy	Multiple-choice test score
Extensive reading for pleasure	55%	90%
Instructional reading	45%	75%
Frustration threshold	35%	50%

The objective was to look at texts that could be considered typical and routinely manageable for students in the courses where the texts were assigned. Naturally, in any course there may be particular readings which are especially difficult and are useful only with extra support to allow students to overcome the texts' inherent difficulty. On the other hand, and probably less commonly, there may be some readings that are especially easy. Such may be the case with natural dialogue or other material not in conventional running text format. It was my intention to avoid both extremes in the sample being analyzed.

The texts varied in length from 211 to 2192 words, averaging 654 for lower division texts and 882 for upper division. Analysis was done on complete passages, excluding titles but including other headings. Footnotes, glossaries, exercises, and any other additions to the main text were excluded. Microsoft Word computed passage length, words per sentence, characters per word, and Flesch Reading Ease and Flesch-Kincaid readability scores for each passage. The Miyazaki EFL Readability Index table was used to look up scaled scores on that measure, and the unscaled version of the Miyazaki formula was used to predict mean cloze scores.

Most of the Miyazaki Index scores for lower division texts were found in a relatively narrow range between 46 and 55, with an inclusive mean of 48.8. Only five texts fell below this range and one above. The 5% trimmed mean, reducing the influence of possible outliers, slightly raised the mean to 49.4. The corresponding predicted mean cloze scores were 23.6% and 23.9%. The passages ranged on the Flesch Reading Ease scale from 33 to 77, with a mean of 59 (trimmed, 60). On the Flesch-Kincaid scale, the range was 6.3 to 12+, (the scale tops out at 12 in MS Word) with a mean of 8.6 (trimmed, 8.5).

The upper division texts were less tightly clustered, ranging from a fairly difficult Miyazaki Index score of 21 (Flesch 32, Flesch-Kincaid 12+) to a middle-range 48 (Flesch 67, Flesch-Kincaid 8.7), with a mean of 35.7 (Flesch 48, Flesch-Kincaid 11). The mean cloze score was computed to be 15.2%. In this group, trimming the mean made almost no difference (possibly because outliers cancelled each other?).

If the assumption is justified that passages in this set were used successfully in the respective courses, we may conclude that first and second year students, on average, can successfully access texts with a readability of about 50 on the Miyazaki Index, or about 60 on the Flesch and 8.5 on the Flesch-Kincaid. On such texts we might expect the average lower division student to achieve a cloze score of about 24% on such a text. Relating this result to the first two studies, that average student would probably report finding such a text fairly difficult at the same time he or she is handling the text in a way that adequately fulfills its intended function in the course.

The findings are not as convenient to accept in the upper division, where we are asked to believe that a mean cloze score of only 15% represents successful

coping with a text, and where it would appear that some texts are still a good deal harder than that. Yet these texts are used and apparently with satisfactory results. The situation begs for an explanation. The following conjectures, while certainly not an explanation, may at least mark out a field for research.

One possibility is, of course, that the students do not cope successfully with the more difficult texts. This is doubtless true of at least some students with at least some texts: some students fail some assignments, and some fail courses. But the large majority do not, and in fact there is plenty of evidence not only that most students do creditable work in their content courses but also that they use these reading materials in doing it.

A second possibility is while students may be using the difficult texts in some way, they are not actually *reading* it in a usual sense of the word. Using a combination of bottom-up decoding and top-down processing and plenty of time, students work the difficult text in much the way I have to work a difficult text by Heidegger or T.S. Eliot. That doesn't mean sitting down and reading straight through at a nice 300 words per minute and answering a comprehension test at the end, much less guessing every 5th word as I read. Yet I *am* able to read Heidegger and Eliot, given enough time to work everything out that I have to in the process. At least I *call* that reading, and perhaps what my students are doing with a difficult text I give them is not very different.

A third possibility is that the relationship of cloze performance to performance on a multiple-choice comprehension test is not the same for EFL readers as for native English readers. Perhaps the EFL student who scores a 25% cloze accuracy on a passage understands that passage better than does a native English speaker who scores only 25%. After all, the data that associate a frustration level difficulty with 35% cloze accuracy are describing native English speakers, and school children at that. Research remains to be done to establish EFL equivalencies between cloze scores and standard comprehension measures. But this kind of research is very difficult to design. In the absence of such information our observation of cloze scores on material that students appear to read for classroom purposes suggests that cloze score/comprehension test equivalencies in EFL will be found to be considerably lower than the traditional ones.

A fourth possibility is that the occasions for academic EFL reading need to be described differently than for native speakers. Recall that the three levels of 55%, 45%, and 35% are said to be criteria for reading for pleasure, instructional reading, and frustration. In fact, how common is it for EFL learners to pick up an English text to read for pleasure apart from the pleasure of improving their English reading skills? Virtually every occasion for reading the EFL text will be instructional, so that the most important distinctions to be made are those regarding the nature and level of support provided and the learning objectives involved.

One final possibility that might help explain why EFL students appear to have a higher tolerance of difficulty than their NS counterparts is that EFL learners naturally expect to have to work harder at reading in a language they are only just learning. They may not like it, but they accept it as a necessary part of their project as EFL learners. Thus their ratings reflect not so much a comfort rating as a rating of what they regard as reasonable and workable for the enterprise. Most probably, the relative weights given to comfort and to reasonableness in the students' ratings is a confounding variable in these studies. However, the fact that the ratings are more generous to the readings than we might expect, based on the difficulty thresholds

traditionally claimed for native speakers, must be in some measure due to this underlying difference in the occasion for reading.

While these studies are too small and some of the important variables too loosely defined to support confident conclusions about specific performance criteria, they call into question the applicability in EFL of native English benchmarks long taken for granted. They also suggest at least tentative guidelines for interpreting readability test results for EFL readers, based on one successful content-based EFL educational program. From the readability scores relative to difficulty ratings which students give, it appears that, on average, our students find the readings we assign them to be a little difficult, but not so difficult that they cannot handle them in a way that we as teachers find satisfactory *for the purposes we intend and with the support we give*.

In this context the Miyazaki Index becomes a meaningful tool for getting a sense of how our students will perceive the readings we give them, how well they may perform with the readings, and how the readings compare to other work being assigned at MIC. Finally, however, we must acknowledge that reading is a complex process, and there are many linguistic, cultural, and individual learner factors that make assessing difficulty as much an art as a science.

We invite others to try the Miyazaki EFL Readability Index. Of course, differences between MIC's educational program and those of other schools may be reflected in different performance criteria associated with readability scores. This may be checked by doing a survey of course readings similar to the one we did in 2000, being careful to examine only texts that have been judged by teachers to have been used successfully with classes. We offer the Miyazaki Index as an appropriate readability measure for such a project.

A Note of Caution

Readability scores are useful in checking our own writing, but they need to be used with care. Readability formulas assume texts that are authentic, that conform to normal usage, and whose vocabulary, grammar, and discourse structures are natural and appropriate for the readers and purposes for which they are being used. Manipulating word or sentence length to improve a readability score without regard to how that may affect cohesion and coherence of the text may inadvertently cancel the reliability of the score and make the reading more difficult, not less.² This caution should not keep us from using readability formulas to check the level of our writing for students. At the same time, however, we need to remember that any teaching tool, however valid, is used best to assist, not replace, the practiced judgment of skilled teachers.

References

- Bormuth, J. (1968). Cloze readability criterion reference scores. *Journal of Educational Measurement*, 5, 189-196.
- Bormuth, J. R. (1971). *Development of standards of readability: Toward a rational criterion of passage performance*. U. S. Department of Health, Education, & Welfare (ERIC Doc. No. ED O54 233).

² For more information on this complex subject, see Leow (1993) and Yano, Long & Ros (1994), both with references to earlier literature.

- Bruce, B., & Rubin, A. (1988). Readability formulas: Matching tool and task. In A. Davison & G. M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered* (pp. 5-22). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Chall, J. & Dale, E. (1995). *Readability revisited: The New Dale-Chall Readability Formula*. Cambridge, MA: Brookline Books.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.
- Greenfield, G. (1999). *Classic readability formulas in an EFL context: Are they valid for Japanese speakers?* Ed.D. dissertation, Temple University. University Microfilms No. 99-38670.
- Greenfield, G. (2001). Performance criteria for EFL/ESL reading difficulty. Paper presented at TESOL 2001 Conference, St. Louis, Missouri, March 2, 2001. Event #5271 (audio taped).
- Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel*. Research Branch Report 8-75. Millington, TN: Naval Technical Training, U. S. Naval Air Station Memphis.
- Leow, R. P. (1993). To simplify or not to simply: A look at intake. *Studies in Second Language Acquisition*, 15, 333-355.
- Tinker, M.A. (1965). *Bases for effective reading*. Minneapolis: University of Minnesota Press.
- Yano, Y., Long, M., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, 44(2), 189-219.