Readability of Texts: Human Evaluation Versus Computer Index

Pooneh Heydari

Department of Foreign Languages,
Shiraz Branch, Islamic Azad University, Shiraz, Iran.

A. Mehdi Riazi

Macquarie University, Australia.

Abstract *This paper reports a study which aimed at exploring if there is any difference between the evaluation of EFL expert readers and computer-based evaluation of English text difficulty. 43 participants including university EFL instructors and graduate students read 10 different English passages and completed a Likert-type scale on their perception of the different components of text difficulty. On the other hand, the same 10 English texts were fed into Word Program and Flesch Readability index of the texts were calculated. Then comparisons were made to see if readers' evaluation of texts were the same or different from the calculated ones. Results of the study revealed significant differences between participants' evaluation of text difficulty and the Flesch Readability index of the texts. Findings also indicated that there was no significant difference between EFL instructors and graduate students' evaluation of the text difficulty. The findings of the study imply that while readability formulas are valuable measures for evaluating level of text difficulty, they should be used cautiously. Further research seems necessary to check the validity of the readability formulas and the findings of the present study.*

Keywords: *text Readability; readability formula; validity of readability formulas; Flesch Reading Ease Readability Formula.*

1. Introduction

How easy or difficult is it to read a text? How clearly does a text express ideas and emotions? These questions are inextricably bound up with the concept of readability (Bailin & Grafstein, 2001). According to Richards, *et al.* (1992, p. 306), readability means: "how easily written materials can be read and understood. This depends on several factors including the average length of sentences, the number of new words contained, and the grammatical complexity of the language used in a passage." Other definitions have been proposed for readability (see, e.g., Dale & Chall, 1949; McLaghlin, 1969).

Procedures used to measure readability are known as readability formulas. Today, there are more than 40 different readability formulas used to measure readability, but some of them are better known and more popular than the others. Perhaps, the most common and the most publicized readability formula was the one credited to Rudolph Flesch (1948). The popularity of his formula made Flesch a leading authority on readability.

Flesch Reading Ease Readability Formula (1948) has also been incorporated and installed in Microsoft Office Word. A text in word can be checked for its spelling and grammar, as well as its readability level. Readability index is based on the average number of syllables per word and words per sentence. Flesch Reading Ease Readability Formula rates texts on a 100-point scale—the higher the score, the easier it is to understand the text. The Flesch Reading Ease Readability Formula is usually defined by the following formula (Flesch, 1948):

$$206.835 - (1.015 \times ASL) - (84.6 \times ASW)$$

Where, ASL is the Average Sentence Length (the number of words divided by the number of sentences), and ASW is the Average of Syllables per Word (the number of syllables divided by the number of words). DuBay (2004) provides interpretation of the Flesch Reading Ease Score as indicated in Table 1.

Table 1: Description and predicted reading grade for Flesch Reading Ease Score (DuBay, 2004)

Reading Ease Score	Description	Predicted Reading Grade	Estimated Percentage of U.S. Adults
0-30	very difficult	college graduate	4.5%
30-40	difficult	college grade	33%
50-60	fairly difficult	10th-12th grade	54%
60-70	standard	8th-9th grade	83%
70-80	fairly easy	7th grade	88%
80-90	easy	6th grade	91%
90-100	very easy	5th grade	93%

It is not clear why the 40-50 range is left out in DuBay's table. The next section presents a review of the background on the text readability.

2. Review of the Readability of the Texts

The first strand of studies on readability present different readability formulas devised to measure the readability of texts (Flesch, 1948; Dale-Chall, 1948 cited in DuBay, 2004; Gunning, 1952; Fry, 1968; McLaughlin, 1969; Flesch-Kincaid Formula, 1975 cited in Greenfield, 1999). In fact, the earliest investigations of readability were conducted by asking students, librarians, and teachers what seemed to make texts readable. Such studies led to the development of mathematical formulas. Today, readability evaluation of the texts is calculated by computer programs. As such, most grammar or editing software programs can determine the readability level of the written materials.

The second group of studies on readability deals with the application of the readability formulas. Readability formulas have had a wide range of applications. Indeed, they were originally created for testing the readability level of school textbooks (Serevin & Tankard, 1992 cited in Balachandran, 1997). Fry (1986, p. 1) pointed out that "articles on the readability formulas are among the most frequently cited articles of all types of educational research." The applications give researchers an objective means for controlling the difficulty of passages in their experiments (DuBay, 2004). Today, readability formulas can be applied to anything from textbooks to government documents and they are more popular than ever. Some of such applications are as follows: (educational system: Kennedy, 1979; Reed, 1988; mass media including newspapers: Lostutter, 1949; Fusaro & Conover, 1983; newsletters: Balachandran, 1997; wire services: Catalano, 1990; brochures: Christ & Pharr, 1980; websites: Baker, Wilson, & Kars, 1997; Graber, Roller, & Kaeble, 1999; manuals: Stahl, Henk, & Eilers, 1995; TV programs: Vancura, 1955; and court actions and legislation: DuBay, 2004). Moreover, there are readability formulas for Spanish, French, Dutch, Swedish, Russian, Hebrew, Hindi, Chinese, Korean (see Rabin, 1988 cited in DuBay, 2004), and Farsi (Dayyani, 1993). The formulas have survived 80 years of intensive application, investigation, and controversy with both their credentials and limitations remaining intact.

The third group of studies discusses the pros and cons views toward the use of readability formulas. Such studies mostly deal with a closer examination of the formulas' underlying principles (Kirkwood & Wolfe, 1980; Bertram & Newman, 1981; Frase, Rubin, Starr, & Plung, 1981; Bailin & Grafstein, 2001), the concept of

their validity and appropriateness for either native or second language learners (Froese, 1971; Hamsik, 1984; Brown, 1998; Greenfield, 1999; Rezaei, 2000; Ardoin, Suldo, Witt, Aldrich, & McDonald, 2005), the discrepancy between the scores of different formulas (Chen, 1986; DuBay, 2004) , and so forth. Such studies help readers to become familiar with the background of the formulas, the theory on which they stand, what they are good for and what they are not.

If any conclusion is possible to draw from the hodge-podge of studies done on readability formulas, it is that there are two opposite views toward the use of them. Both of these two views have been advocated by different researchers and there is enough empirical evidence for each to be true. Thus, it can be declared openly that the formulas have both advantages and disadvantages.

Advantages of using readability formulas:

1. By definition, readability formulas measure the grade-level readers must have to read a given text. The results from using readability formulas provide the writer of the text with much needed information to reach his target audience.
2. Readability formulas do not require the readers to first go through the text to decide if the text is too hard or too easy to read. By readability formulas, one can know ahead of time if his readers can understand the material. This can save time, money and energy.
3. Readability formulas are text-based formulas; many researchers and readers find them easy to use.
4. Today, readability formulas can be performed by computer. As such, most grammar or editing software today can determine the readability level of written materials.
5. Readability formulas help writers convert their written material into plain language.

Disadvantages of using readability formulas:

1. Unfortunately, readability formulas are not of much help if one wants to know how well the target audience understands the text.
2. Due to many readability formulas, there is an increasing chance of getting wide variation in results of a same text.
3. Readability formulas cannot measure the context, prior knowledge, interest level, difficulty of concept, or coherence of text.

Indeed, it is a particular propitious moment to re-examine the use of readability formulas as a measure of reading difficulty. This abundance of research by itself is the material proof for the significance of the topic under study. Of course, this fact necessitates further research in this area, and the present study was set up to this end.

3. The Objectives

This study aimed at helping EFL educators and practitioners to make more objective decisions about how to go about selecting, revising, teaching, and evaluating EFL texts by evaluating the readability formulas and correlating standard indexes with the indices obtained by several other means like students' self-assessment of texts and expertise rating. Revisiting the concept of EFL text-readability formulas, the consequent related issues in the EFL reading skill such as reliability and validity, plus the scarcity of research in this area for EFL learners constitute the significance of the scope of this study.

The following questions are to be answered through this study:

1. Is there any difference between EFL expert readers' evaluation of text-readability and computer-based readability index?
2. Is there any difference between the two groups of expert readers' (EFL instructors and EFL graduate students) evaluation of text-readability?

4. Method

4.1 .Participants

This study was conducted with an overall number of 43 participants. Eight of these participants were male EFL professors and instructors at a major university in Iran. The other 35 were selected from among graduate students studying different EFL-related majors at two universities. The participants were chosen using convenient sampling procedure.

4.2. Materials and Procedures

First, 10 reading passages were selected from the book of *Patterns: A Short Prose Reader, 5th Edition* (Conlin, 1998), a reading text-book for EFL students studying at the advanced level of reading/writing courses locally. This book contains 95 passages collected from authentic samples of American prose from a wide and diverse pool of writers including statesmen, professional authors, educators, and practitioners. The 10 chosen passages had different characteristics in terms of text-content, type, genre, organization, and readability index—the factors which are internal to reading texts. They also had different methods of development including cause and effect, comparison and contrast, description, process, classification and division, narration, and definition.

Before the study and as a pilot study, a number of readers similar to participants of the study were asked to read all of the 10 selected passages. Based on the feedback obtained from them, it was found that the number of texts was demanding and time consuming for the participants. Therefore, based on the Flesch Rank Order, the 10 passages were divided up into two sets; each set containing five passages of different readability index. The two sets were relatively equal in terms of readability index; 56.1 and 56.6, respectively. Table 2 presents the texts in each set along with Flesch score and rank order for each of the texts.

Table 2: Readability Score and Rank Order of the Passages Using Flesch Reading Ease Readability Formula

Passage	Flesch Score	Rank Order	Set
1	31.2	8	A
2	73.7	3	B
3	63.7	4	A
4	63.7	4	A
5	28.71	9	B
6	53.9	5	B
7	78.2	1	B
8	46.4	7	A
9	48.6	6	B
10	75.5	2	A

4.3. Instrument

A questionnaire aimed at tapping responses from the participants on the elements making up text-readability of the 10 passages using Likert-scale was devised (see Appendix B). Reviewing the literature on readability, 11 most effective factors operating on reading difficulty were chosen as the items of the questionnaire. Different experts have taken these eleven factors into account as active elements in assessing reading difficulty of the texts.

Since the items have been suggested by reputable authorities of the field, the content validity of the questionnaire is taken for granted. The factors are as follows:

1. Word familiarity: Flesch (1948), Dale & Chall (1948 cited in DuBay, 2004), Gunning (1952), Fry (1968), McLaughlin (1969), Bailin & Grafstein (2001),
2. Word frequency: Thorndike (1921), Lively & Pressey (1923), Klare (1968), Chall & Dale (1995), Graesser *et al.* (2004),
3. Topic and content familiarity: Chang (2006), Pulido (2007), Combs (2008),
4. Sentence length: Kitson (1921), Flesch (1948), Dale & Chall (1948 cited in DuBay, 2004), Gunning (1952), Fry (1968), McLaughlin (1969), Catalano (1990),
5. Pronoun density: Graesser *et al.* (2004),
6. Number of ambiguous words: Padak (1993), Graesser *et al.* (2004),
7. Syntactic complexity: Padak (1993), Bailin & Grafstein (2001), Graesser *et al.* (2004),
8. Concreteness: Graesser *et al.* (2004),
9. Imageability: Graesser *et al.* (2004),
10. Concept clarity: Bertram & Newman (1981), Padak (1993), Graesser *et al.* (2004), and
11. Time: DuBay (2004)

Attempts were also made to ensure that the items are understandable to the participants. The reliability of the questionnaire using the data from the participants was calculated using internal consistency through Cronbach's alpha and was found to be 0.78.

4.4. Data Collection Procedures

For the first part of the study, the participants were randomly divided into two groups: the first group including four experts (EFL professors) and 18 graduate students who received the first set of passages (Set A), and the second group including four experts and 17 graduate students received the second set (Set B). Participants were asked to provide feedback on text characteristics, particularly text-readability, difficulty rank-order, and the like by filling out the questionnaire. The completed questionnaires were collected for analysis. On the other hand, all the 10 texts were fed into Microsoft Office Word and the readability level of the passages were calculated.

4.5. Data Analysis Procedures

The data were analyzed through appropriate statistical procedures including one-sample t-test and Mann-Whitney U test.

For the first research question of the study investigating if there was any difference between EFL expert reades evaluation of text-readability level and computer-based readability index, a one-sample t-test was run. Also, to find out if there was any difference between EFL experts' and EFL students' evaluation of text-readability, the non-parametric Mann-Whitney U-test was used which is the equivalent statistical test for the parametric independent sample t-test. Mann-Whitney U-test was used since the number of participants was low.

5. Results and Discussion

Table 3 presents the descriptive statistics of the two sets of passages as evaluated by the participants including both experts and graduate students and the readability indices of the texts as presented in Table 2.

Table 3. Descriptive statistics of the texts based on Flesch and participants' evaluation

Texts	Mean (participants)	SD	Mean (Flesch RI)
(Set A)1	56	15.5	31.2
3	87	16	63.7
4	66	16	63.7
8	56	17	46.4
10	86	17	75.5
(Set B)2	57	19.5	73.7
5	61.5	17.2	28.71
6	60	20.6	53.9
7	72	18.9	78.2
9	61	16.2	48.6

Comparing the means, one can conclude that just in the case of passages 4, 6, and 7 they are close to each other. However, they turned out to be quite far from each other in the case of other passages. To see if such differences were significant or not, a t-test was run on each case. Table 4 shows the results of the comparison between different evaluations.

Table 4. Results of the comparison between different evaluations

Text	Mean (experts)	Mean (Flesch RI)	t value	df	Sig. (2-tailed)
1	56	31.2	7.476	21	.000
2	57	73.7	-3.867	20	.001
3	87	63.7	6.800	21	.000
4	66	63.7	.661	21	.516
5	61.5	28.71	8.730	20	.000
6	60	53.9	1.348	20	.193
7	72	78.2	-1.512	20	.146
8	56	46.4	2.668	21	.014
9	61	48.6	3.595	20	.002
10	86	75.5	2.882	21	.009

As this table shows, in 70% of the cases the differences between the two evaluations are significant. Only three passages—4, 6, and 7—evaluations did not show any significant differences. In other words, teachers' and students' evaluation of text-readability level and the computer-based evaluation of text-readability through the use of Flesch Reading Ease Readability Formula were significantly different for seven passages.

For the second research question of the study, investigating if there was any difference between experts' and postgraduate EFL students' evaluation of text-readability, a Mann-Whitney U test was run. Tables 5 and 6 present the descriptive statistics for the two sets of evaluations.

Table 5: Descriptive statistics for Set A

Text	Participant		N	Mean Rank	Sum of Ranks
1	Expert	1.00	4	14.00	56.00
	Students	2.00	18	10.94	197.00
	Total		22		
3	Expert	1.00	4	7.50	30.00
	Students	2.00	18	12.39	223.00
	Total		22		
4	Expert	1.00	4	9.38	37.50
	Students	2.00	18	11.97	215.50
	Total		22		
8	Expert	1.00	4	11.63	46.50
	Students	2.00	18	11.47	206.50
	Total		22		
10	Expert	1.00	4	8.50	34.00
	Students	2.00	18	12.17	219.00
	Total		22		

Table 6. Descriptive statistics of Set B

Text	Participant		N	Mean Rank	Sum of Ranks
2	Expert	1.00	4	12.50	50.00
	Students	2.00	17	10.65	181.00
	Total		21		
5	Expert	1.00	4	12.63	50.50
	Students	2.00	17	10.62	180.50
	Total		21		
6	Expert	1.00	4	11.75	47.00
	Students	2.00	17	10.82	184.00
	Total		21		
7	Expert	1.00	4	8.25	33.00
	Students	2.00	17	11.65	198.00
	Total		21		
9	Expert	1.00	4	13.75	55.00
	Students	2.00	17	10.35	176.00
	Total		21		

By virtue of the above tables, it is clear that the mean ranks of each passage obtained from the experts' and students' evaluations are quite close to each other. Tables 7 and 8 show the results of the comparison. Tables 7 and 8 present the results of Mann-Whitney U-tests.

Table 7. Mann-Whitney U test for Set A

	Text1	Text3	Text4	Text8	Text10
Mann-Whitney U	26.000	20.000	27.500	35.500	24.000
Wilcoxon W	197.000	30.000	37.500	206.500	34.000
Z	-.851	-1.367	-.724	-.043	-1.022
Asymp. Sig. (2-tailed)	.395	.172	.469	.966	.307
Exact Sig. [2*(1-tailed Sig.)]	.434(a)	.195(a)	.484(a)	.967(a)	.342(a)

Table 8. Mann-Whitney U test for Set B

	Text 2	Text 5	Text 6	Text 7	Text 9
Mann-Whitney U	28.000	27.500	31.000	23.000	23.000
Wilcoxon W	181.000	180.500	184.000	33.000	176.000
Z	-.537	-.582	-.269	-.985	-.986
Asymp. Sig. (2-tailed)	.591	.560	.788	.325	.324
Exact Sig. [2*(1-tailed Sig.)]	.635(a)	.574(a)	.829(a)	.362(a)	.362(a)

It can be seen that none of the significance levels was smaller than .05, so it is concluded that experts' and EFL students' evaluations did not differ significantly in terms of their evaluation of text-readability level.

The first part of the current study's results is consistent with those of Froese (1971), Carrell (1987), Brown (1998), Rezaei (2000), Ardoin, Suldo, Witt, Aldrich, and McDonald (2005), who attempted to re-evaluate the validity of different readability formulas. These researchers studied the validity of readability formulas in comparison to various independent criteria of reading difficulty such as cloze units, multiple-choice items, and equivalent forms. All of them asserted that the formulas are not valid measures of the difficulty of the written materials. Findings of the present study are in sharp contrast to Hamsik (1984), Fry (1989), and Greenfields' (1999) who asserted that not only readability formulas can prove to be valuable tools for measuring the difficulty of texts for native English speakers, but also they do measure the readability level of texts for EFL learners.

The second goal of this study was to explore the differences between teachers' and students' evaluation of text difficulty. As Tables 5 and 6 indicate, the mean ranks obtained for each passage from different evaluations are quite close to each other and did not show to be significantly different. This finding is not, of course, in line with Montgomery's (1985) findings, who reported no significant relationship between students' and teachers' judgment of text difficulty.

6.Conclusions

This study was carried out to investigate the validity of the Flesch Reading Ease Readability Formula. Firstly, the study investigated if there is any significant difference between instructors' and students' evaluation of

text difficulty and computer-based readability index which is based on Flesch formula. Results of t-test suggested that there was a significant difference between the two indices for seven passages while no significant results were found for three passages. The following table summarizes results of the first part of the study.

Table 9. Summary of the findings of the first part of the study

	Text						
	1	2	3	5	8	9	10
Participants' evaluation of the text difficulty	56	57	87	61.5	56	61	86
Level of difficulty	easy	easy	very easy	easy	easy	easy	very easy
Readability index based on Flesch formula	31.2	73.7	63.7	28.71	46.4	48.6	75.5
Level of readability	difficult	fairly easy	standard	very difficult	--	--	fairly easy

As can be seen in Table 9, except for texts 2 and 10 which results might approximate each other, the other five texts were evaluated differently by participants and by the Flesch formula.

On the other hand, since there were two groups of participants, namely, university instructors and graduate students, we were interested to find out if the evaluation of these two groups in respect to the text difficulty were different. As indicated in Tables 7 and 8 no significant difference was found meaning that despite some differences between the two groups of the participants in terms of their reading English text ability, their evaluation of text difficulty was not significantly different from each other.

The findings of the study have implications for materials and test developers. While further research is needed to validate the findings of the present study, it can be suggested that decisions for text selection in terms of text difficulty and readability are not based on readability formula per se. Human judgments are needed to be considered as well.

References

Ardoin, S. P., Suldo, S. M., Witt, J., Ardrich, S., & McDonald, E. (2005). Accuracy of readability estimates' predictions of CBM performance [Electronic version]. *School Psychology Quarterly, 20*(1), 1-22. Retrieved December 23, 2007, from http://proquest.umi.com/pqdweb?did=818687841&sid=2&Fmt=&clientId=46449&RQT=309&VName=PQD

Bailin, A., & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language & Communication, 21*, 285-301.

Baker, L. M., Wilson, F. L., & Kars, M. (1997). The readability of medical information on Info Trac: Does it meet the needs of people with low literacy skills? *References & Users Services Quarterly, 37*(2), 155-160.

Balachandran, B. (1997). *Readability standards of newsletters.* Unpublished master's thesis, University of California, Fresno, U.S.

Bertram, B., & Newman, S. (1981). *Why readability formulas fail* (Report No. 28). Illinois University, Urbana: Center for the Study of Reading. (EricDocument Service No. ED205915)

Brown, J. D. (1998). An EFL readability index. *JALT Journal, 20*, 7-36.

Carrell, P. (1987). Readability in ESL. *Reading in a Foreign Language, 4*, 21-40.

Catalano, K. (1990). On the wire: How six news services are exceeding readability standards. *Journalism Quarterly, 67*(1), 97-103.

Chall, J. S., & Dale, E. (1995). *Readability revisited, the new Dale-Chall readability formula.* Cambridge, MA: Bookline Books.

Chang, C. (2006). Effects of topic familiarity and linguistic difficulty on the reading strategies and mental representations of nonnative readers of Chinese. *Journal of Language and Learning, 4*(2), 172-197.

Chen, W. S. (1986). A comparison of seven computerized readability formulas as applied to elementary social studies textbooks. Unpublished doctoral dissertation, Columbia University College, U.S. Retrieved November 10, 2008, from: http://proquest.umi.com/pqdweb?did=771892981&sid=9&Fmt=4&clientl   d=46449&RQT=309&VName=PQD

Combs, C. (2008). Topic familiarity and input enhancement: An empirical investigation. *Applied Linguistics, 8*(2), 1-51.

Conlin, M. L. (1998). *Patterns: A short prose reader* (5th ed.). Boston, MA: Houghton Mofflin Company.

Christ, W. G., & Pharr, P. (1980). Readability of brochures produced by state of Florida. *Journalism Quarterly, 57*, 150-159.

Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English, 26*, 23-33.

Dayyani, M. H. (1993). An assessment of the readability of special Farsi texts written for newly-literates. *Adult Education and Development, 2*(3), 27-48.

DuBay, W. H. (2004). The principles of readability. Retrieved February 12, 2008, from: http://www.nald.ca/fulltext/readab/readab.pdf

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221-233.

Frase, L. T., Rubin, A., Starr, K., & Plung, D. L. (1981). Readability formulas: Used or abused? [Electronic version]. *IEEE Transactions on Professional Communication, PC24* (1), 48-54. Retrieved December 23, 2007, from: http://proquest.umi.com/pqdweb?did=1135513&sid=9&Ftm=2&clientId=   46449&RQT=309&VName=PQD

Froese, V. (1971). Cloze readability versus the Dale-Chall formula. Paper presented at the meeting of the International Reading Association, Atlantic City, NJ.

Fry, E. B. (1968). A readability formula that saves time. *Journal of Reading, 11*, 513-516.

Fry, E. B. (1986). *Varied uses of readability measurement.* Paper presented at the 31st Annual Meeting of the International Reading Association, Philadelphia, PA.

Fry, E. B. (1989). Reading formulas – maligned but valid. *Journal of Reading, 32* (4), 292-297.

Fusaro, J. A., & Conover, W. M. (1983). Readability of two tabloid and two nontabloid papers. *Journalism Quarterly, 50*, 360-363.

Graber, M. A., Roller, C. M., & Kaeble, B. (1999). Readability levels of patient education material on the Word Wide Web. *The Journal of Family   Practice, 48*(1), 58-61.

Graesser, G. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers, 36*, 193-202.

Greenfield, G. R. (1999). *Classic readability formulas in an EFL context: Are they valid for Japanese speakers?* Unpublished doctoral dissertation, University of Temple, U.S.

Gunning, R. (1952). *The technique of clear writing.* New York: McGraw-Hill.

Hamsik, M. J. (1984). *Reading, readability, and the ESL reader.* Unpublished doctoral dissertation, The Florida University, U.S. Retrieved December 23, 2007, from: http://proquest.umi.com/pqdweb?did=749006291&sid=1&ftm=2&clientId =46449&RQT=309&VName=PQD

Kennedy, K. (1979). The reading levels of high school physics texts. *The Physics Teacher, 17*, 165-167.

Kitson, H. D. (1921). *The mind of the buyer.* New York: Macmillan.

Kirkwood, K. J., & Wolfe, R. G. (1980). *Matching students and reading materials: A cloze-procedure method for assessing the reading ability of students and the readability of textual materials.* Toronto: Ontario Department of Education. (ERIC Document Reproduction in Service No. ED 195 928)

Klare, G. R. (1963). *The measurement of readability.* Ames, IA: Iowa State University Press.

Klare, G. R. (1968). The role of the word frequency in readability. *Elementary English, 45*, 12-22.

Lively, B. A., & Pressey, S. L. (1923). A method for measuring the vocabulary burden of textbooks. *Educational Administration and Supervision IncludingTeacher Training, 9*, 389-398.

Lostutter, M. (1949). Some critical factors of newspaper readability. *Journal Quarterly, 26*, 307-314.

McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading, 12*(8), 639-646.

Mehrpour, S., & Riazi, A. M. (2004). The impact of text length on EFL students' reading comprehension. *Asian EFL Journal, 6*(3), 1-13.

Montgomery, L. E. (1985). *A study of the relationships of readability among elementary science textbooks adopted by Texas using five measures (Cloze, Fry, Dale-Chall).* Unpublished doctoral dissertation, Woman's   University, Texas. Retrieved November 10, 2008, from: http://proquest.umi.com/pqweb?did=753639141&sid=2&Fmt=2&clientId   =46449&RQT=309

Padak, N. (1993). Looking at readability for adult literacy learners. Retrieved December 10, 2008, from: http://literacy.kent.edu/Oasis/Pubs/0300-45.pdf

Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical differencing and retention through reading. *Applied Linguistics,28*(1), 66-86.

Reed, K. X. (1988). *An analysis of reading levels of students and readability levels of textbooks at second junior colleges in the state of Alabama.* Unpublished doctoral dissertation, University of Auburn. Retrieved October 12, 2008, from: http://proquest.umi.com/pdqweb?did=744752841&sid=2&Fmt=2&clientl d=46449&PQT=309&VName=PQD

Rezaei, A. A. (2000). The validity of the "Fog-Index of Readability". *Journal of Humanities of Islamic Republic of Iran, 7*(4), 17-27.

Richards, J. C., Platt, J., & Platt, H. (1992). *Longman dictionary of language teaching and applied linguistics.* London: Longman.

Stahl, N. A., Henk, W. A., Eilers, U. (1995). Are drivers' manuals understandable? [Electronic version]. *Transportation Quarterly, 49*(1), 105. Retrieved December 10, 2008, from: http://proquest.umi.com/pqdweb?did=4468887&sid=2&Fmt=2&clientId =46449&RQT=3098VName=PQD

Thorndike, E. L. (1921). *The teacher's workbook.* New York, NY: Teacher's College, Columbia University.

Vancura, R. H. (1995). Flesch readability formula applied to television programs. *Journal of Applied Psychology, 39*(1), 47-48.

Appendices

Appendix A

Set A

Passage 1:

Cause and Effect
The Power of Place

Few places are as overly stimulating as the rush-hour subway, yet some of the less overtly jittery spots where we spend far more time exact their tolls as well. As we move further into the postmodern age of information, the workplace is changing fast, causing occupational safety specialists to focus on problems that would have seemed light-weight to their predecessors. Not long ago, for example, "industrial fatigue" meant the hard-hat exhaustion of steelworkers and coal miners. Now it is just as likely to refer to the weariness, eyestrain, and aches and pains of computer operators who spend long periods with poorly designed VDTs (video display terminals or computer screens), desks, and chairs. Environmental psychologists have shown that the proper adjustment of a single element at a computer station calls for pains-taking microanalysis. To evaluate the screen, for example, one must consider its height, tilt, and distance from the operator, the size and clarity of its characters, and its brightness, glare, and flickering. Because its effects are often subtle, combined with other irritants, and bother us after exposure ceases, trying to pinpoint an environmental stressor is difficult for the layman. As a result, we often end up blaming its noxious influence on something else — the project, the boss, or "stress" in general — thus perpetuating the dilemma.

Passage 3:

Description
The Quiet Odyssey

In Los Angeles in 1950, we found many minority women working in sewing factories making garments of every sort of fifty cents an hour, eight hours a day. After several years, the wage went up to one dollar an hour. The sewing rooms were dirty and very dusty, with lint and dust filling the air like fog. The rooms had no air conditioning and no windows. The dust settling on the heads of women made their hair look gray by the end of the day. The loud power-driven sewing machines working at full speed all at once made a thundering noise that deafened the ear. It was a frightful thing to listen to for eight hours every weekday. I tried it once for several months; the experience made me admire all those women who endured it for years in order to send their children to colleges and universities. I have seen those children return home as doctors, lawyers, and engineers, thus rewarding their parents for their sacrifices.

Passage 4:

Process
Like Water for Chocolate

At the entrance to the ranch Nicholas and Rosalio, in fancy charro costumes, were collecting invitations from the guests as they were arriving. The invitations were beautiful. Alex and Esperanza had prepared them personally. The paper used for the invitations, the black ink used to write them, the gold tint used on the edges of the envelopes, and the wax used to seal them — all those were their pride and joy. Everything had been prepared the traditional way, using the De la Garza family recipes. But they hadn't needed to prepare the black ink, for enough remained from the ink that had been made for Pedro and Rosaura's wedding. It was dried ink; all that had to be done was to add a little water and it was as good as new. The ink is made by mixing eight ounces of gum Arabic, five and a half ounces of gall, four ounces of iron sulfate, two and a half ounces of logwood, and half an ounce of copper sulfate. To make the gold tint used on the edges of the envelopes, take an ounce of orpiment and an ounce of rock crystal, finely ground. Stir these powders into five or six well-beaten egg whites until the mixture is like water. And finally, the sealing wax is made by melting a pound of gum Arabic, half a pound of benzoin, half a pound of calafonia, and a pound of vermilion.

Passage 8:

Classification and Division
Fans

The fight crowd is a beast that lurks in the darkness behind the fringe of white light shed over the first six rows by the incandescent atop the ring, and is not to be trusted with pop bottles or other hardware. The tennis crowd is always preening and shushing itself. The golf crowd is the most unwieldy and most sympathetic, and is the only horde given to mass production of that absurd noise written generally as "tsk tsk tsk tsk," and made between tongue and teeth with head-waggings to denote extreme commiseration. The baseball crowd is the most hysterical, the football crowd the best-natured and the polo crowd the most aristocratic. Racing crowd are the most restless, wrestling crowd the most tolerant, and soccer crowd the most easily incitable to riot and disorder. Every sports crowd takes on the characteristics of the individuals who compose it. Each has its particular note of hysteria, its own little cruelties, mannerism, and bad mannerism, its own code of sportsmanship and its own method of expressing its emotions.

Passage 10:

Narration 2
Freedom

Two years ago, I attempt to escape from mainland China to Hong Kong. I planed and prepared well. I dressed up like a farmer and walked for two days from my village to the border between China and Hong Kong. That night, I was very excited and nerves, but I tried to keep calm. At the boarder there were a lot of sentries who tried to catch people like me, so I put some mud on myself to avoid being noticed. It was not easy for me to pass through the sentries, but I bit my tongue and climbed across the swampy area. Finally, I reached the river that runs across the boarder. I plunged into it. It was icy cold, and I used all my strength to swim as fast as I could. In about twenty minutes, I touched land. I had made it! My happiness was beyond description. But when I stood up, a Hong Kong policeman was immediately beside me. My dream was shattered. I was taken to a police station to wait for a trunk that takes unsuccessful refugees back to China. The police took me in the trunk with a great many other people and we were driven like a herd of buffalo back to China. I had lost my freedom again.

Set B

Passage 2:

Comparison and Contrast
That Learn and Hungry Look

Some people say the business about the jolly fat person is a myth, that all of us chubbies are neurotic, sick, sad people. I disagree. Fat people may not be chortling all day long, but they're a hell of a lot nicer than the wizened and shriveled. Thin people turn surely, mean and hard at a young age because they never learn the value of a hot fudge sundae for easing tension. Thin people don't like gooey soft things because they themselves are neither gooey nor soft. They are crunchy and dull, like carrots. They go straight to the heart of the matter while fat people let things stay all blurry and hazy and vague, the way things actually are. Thin people want to face the truth. Fat people know there is no truth. One of my thin friends is always staring at complex, unsolvable problems and saying, "The key thing is …" Fat people never say that. They know there isn't any such thing as the key thing about anything.

Passage 5:

Classification and Division 1
From Cakewalks to Concert Halls

Ragtime's complex historical legacy was perhaps a major reason for its widespread appeal among both blacks and whites. First and foremost, it was a dance music which drew on both European and African traditions. Second, ragtime was a style grounded in an ongoing, cross-cultural, racial parody: the slaves' parody of their masters, blackface minstrels' trope of the slaves' parody, black minstrels' trope of the blackface parody, and so on. In addition, ragtime was a rural folk music transposed to an urban and industrial context, where its machine-like rhythms became an expression of the lost innocence of bygone days and ways. And finally, as a novel

popular music created by the first generation of African Americans born after slavery, ragtime represented an affirmation of their newly experienced freedoms and an optimistic vision of the future.

Passage 6:

Comparison and Contrast 2
Jungle and Desert

The way of the desert and the way of the jungle represent the two opposite methods of reaching stability at two extremes of density. In the jungle there is plenty of everything life needs except mere space, and it is not for the want of anything else that individuals die or that races have any limit set to their proliferation. Everything is on top of everything else; there is no cranny which is not both occupied and disputed. At every moment, war to the death rages fiercely. The place left vacant by any creature that dies is seized almost instantly by another, and life seems to suffer from nothing except too favorable an environment. In the desert on the other hand, it is the environment itself which serves as the limiting factor. To some extent the struggle of creature against creature is mitigated, though it is of course not abolished even in the vegetable kingdom. For the plant which in the one place would be strangled to death by its neighbor dies a thirsty seeding in the desert because that same neighbor has drawn the scant moisture from the spot of earth out of which it was attempting to spring.

Passage 7:

Narration 1
Vital Signs

Within Janine's playroom, we were some of us handicapped, but none disabled, and in time we were each taught to prove this for ourselves. While I poured the flour for new play dough, Janine asked me about my kindergarten teacher: What she had looked liked with an eye patch, and if she was missing my same eye. What were the hard parts, Janine said, for a teacher like that? Did I think it was sad for her to miss school sometimes, and did she talk about the hospital? What color was her hair, what sort was her eye path, and did I remember if she was pretty? What would I be, Janine said, when I was that age and these surgeries were past? Over the wet salt smell of green dough, I wished to be a doctor with one blue eye, who could talk like this to the sick, who could tell them they were still real. And with her feel for when to stop talking, Janine turned and left me, searching out volunteers to stir up new clay.

Passage 9:

Definition
The Ultimate Kitchen Gadget

It is the ultimate kitchen gadget. It serves as a juicer for lemons, oranges, and grapefruit, and as a combination seed remover and pulp crusher for tomatoes. It functions as a bowel scraper, an egg separator and a remover of unwelcome particles — the stray bit of eggshell, the odd grain of black rice — from mixing bowel or saucepan. It is a thermometer capable of gauging temperatures up to 500 degrees Fahrenheit and, in addition, is a measuring device for dry ingredients in amounts from 1 tablespoon down to 1/8 teaspoon or less, and for whatever liquids may be called for in the cooking of grains and stocks. It can be used as tongs for removing hot cup custards from the oven, as a mixer of water into pastry dough and as a kneader of bread. Best of all, it cleans up in a trice, presents no storage problems, will not chip, rust or tarnish and, if it cannot be said to be unlosable or indestructible, it nevertheless comes with a lifetime guarantee to remain the one household convenience you will have the least desire either to love or to destroy. It is, of course, the human hand.

Appendix B

Give each passage a grade of 0 to 100 in return for each variable.
- 75-100 is considered very easy,
- 50-75 is considered easy,
- 25-50 is considered difficult, and
- 0-25 is considered very difficult.

Passage No.	1 2 3 4 5 6 7 8 9 10
Variables	
Word familiarity (use of short, simple, and familiar words)	
Word frequency (how frequently a word appears in print)	
Topic and content familiarity (background schemata)	
Sentence length	
Density of pronoun (No. of first, second and third pronoun)	
No. of ambiguous words (polysemy)	
Syntactic complexity of sentences (use of simple sentences, active voice, and present tense)	
Concreteness (how concrete or non-abstract the words are)	
Imageability (how easy is it to construct a mental image of the words)	
Concept clarity (semantic explicitness)	
Time (reading time for each passage)	