

USING CLOZE PASSAGES TO ESTIMATE READABILITY FOR RUSSIAN UNIVERSITY STUDENTS: PRELIMINARY STUDY

JAMES DEAN BROWN, GERRIET JANSSEN, JONATHAN TRACE¹, LIUDMILA KOZHEVNIKOVA²

Abstract

This preliminary study examines the relationships between each of six first language (L1) readability indexes and the cloze passage mean performances of Russian EFL students. The cloze passages were created by randomly selecting 50 text passages from an American public library and deleting every 12th word in each passage to create a 30-item cloze procedure. The participants were 5170 EFL students from 38 universities in the Russian Federation. Each student was randomly assigned to take one of the 30-item cloze passages. The L1 readability indexes calculated for each of the 50 passages were the Flesch, Flesch-Kincaid, Fry, Gunning, Fog, and modified Gunning-Fog indexes. The preliminary results indicate that the L1 readability indexes were moderately to highly correlated with each other, but only somewhat correlated with the mean performances of Russian university students on cloze versions of those same passages. These results are discussed in terms of why the L1 readability indexes are moderately to highly correlated with each other but only somewhat correlated to the Russian EFL means. The authors also explain what they are planning in terms of further linguistic analyses (e.g., of variables like average word length, percent of function words, number of syllables per sentence, number of words per paragraph, frequencies of words in the passages, and so forth) and statistical analyses (including at least factor analysis, multiple regression analyses, and structural equation modeling) of these data.

Key Words: second language readability, English language teaching.

INTRODUCTION

First Language Readability

Readability is a concept that describes the degree to which a text is easy or difficult to read. A *readability index* is a numerical scale that estimates the readability or degree reading difficulty that native speakers are likely to have in reading a particular text. For example, the Fry (1977) readability scale is expressed in grade levels for students in the United States ranging from 1 (first grade) to 17+ (graduate school and beyond). Thus a passage with a Fry scale index of 3.5, would be fairly easy because it would be appropriate for children who are native speakers of English in the second half of third grade (or about 8 years old), whereas a passage with an index of 13 would be more difficult because it would be suitable for first-year university-level native speakers of English.

The findings from one study (Brown, Chen, & Wang, 1984) led the first author believe that such L1 readability indexes might be useful indicators of relative passage difficulty in EFL settings. Brown, Chen, and Wang studied the readability of the cards in Stanford Research Associates (SRA) classroom reading kits. Those kits have cards at different grade levels (coded by color) that had previously been established by research into the actual performances on those cards of North American children. Brown, Chen, and Wang started by calculating the Fry readability index for each of the SRA cards. They then compared the resulting Fry scale indexes with the actual native-speaker grade-level performances.

¹ University of Hawai'i at Mānoa

² Samara University

Table 1. *The Accuracy of L1 Fry Readability Estimates (Adapted from Brown, Wang, & Chen, 1984)*

SRA Kit	Passage Grades Based on Student Performance				
		Mean	SD	Min	Max
3A	3.5	3.22	1.20	2	6
	4	4.56	1.42	3	6
	4.5	5.56	0.88	4	7
	5	6.44	0.73	5	7
	6	7.11	0.93	6	8
	7	8.22	2.17	6	13
	8	8.67	1.50	6	10
	9	9.56	1.67	6	12
	10	10.22	1.48	7	12
	11	10.11	2.15	6	12
	4A	8	8.56	1.13	6
9		9.44	0.88	8	10
10		10.44	1.74	9	14
11		11.11	1.83	7	13
12		12.56	1.51	11	16
13		13.11	3.30	9	17+
14		13.25	1.98	9	15

Table 1 shows results for the 3A and 4A SRA kits separately. The grade levels shown in column two were established by research into the performances of L1 native speakers (each grade level consisted of 12 to 14 cards). The mean¹, standard deviation² (*SD*), minimum³ (*Min*), and maximum⁴ (*Max*) for the Fry scale readability estimates for each grade or half-grade level are shown in the last four columns. Notice, in the third column, that the mean grade levels for the Fry index are remarkably close to the actual grade levels of the cards as established by student performance. Clearly, this study demonstrates that there is a strong relationship between the mean grade levels estimated by the Fry scale and the grade levels established on the basis of native-speaker performances.

A large number of English as a first language (L1) readability indexes have been invented over the past 60 years. Chall (1958), Klare (1963; 1984), Zakaluk and Samuels (1988), or Zamanian and Heydari (2012) all provide overviews of the first language readability literature. The L1 readability indexes examined in the current paper are: the Flesch reading ease formula (Flesch, 1948), the Flesch-Kincaid readability index (as described in Klare, 1984), the Fry readability scale (see Fry, 1985), as well as the Gunning, Fog, and modified Gunning-Fog readability indexes (see Larson, 1987).

The simplest way to explain the *L1 readability indexes* is to show the equations that define them. For example, Flesch's (1948) equation multiplies the average number of syllables per word in the text by .846, then subtracts the result from 206.835. From this result, the equation subtracts 1.015 times the average number of words per sentence. The actual equation for the Flesch reading ease index is:

1. Flesch Reading Ease Formula (Flesch, 1948)

$$= 206.835 - .846(\text{syllables} / \text{words}) - 1.015(\text{words} / \text{sentences})$$

The other indexes described below (numbers 2 to 6) are similar manipulations of the numbers of syllables, words, long words, easy words, hard words, sentences, etc. For those who are interested, we also provide a seminal reference for each equation.

2. Flesch-Kincaid Index (as cited in Klare, 1984)

$$= .39(\text{words} / \text{sentences}) + 11.8(\text{syllables} / \text{words}) - 15.59$$

3. Fry Scale (Fry, 1977 or 1985)

= on the Fry reading graph (see Fry, 1985), the grade value at the point where the coordinates for sentences per 100 words and syllables per 100 words cross

¹ Here the *mean* can be interpreted as the more familiar arithmetic *average*.

² The *standard deviation* is a sort of average of the distances from the mean of all the values in the data; as such, it is an indicator of how much the values are dispersed around the mean.

³ The *minimum* is the lowest value in the set of numbers.

⁴ The *maximum* is the highest value in the set of numbers.

4. Gunning Index (as cited in Carrell, 1987)

$$= .40(\text{words} / \text{sentences} + \% \text{ of words over two syllables})$$

5. Fog Count (as cited in Carrell, 1987)

$$= \frac{\left\{ \frac{\text{easy words} + 3(\text{hard words})}{\text{sentences}} - 3 \right\}}{2}$$

6. Gunning-Fog Index (Larson, 1987)

$$= \left\{ \frac{\text{words}}{\text{sentences}} + 100 \right\} \times \left\{ \frac{\text{long words}}{\text{sentences}} \frac{\text{sentences}}{\text{words}} \right\}$$

Second Language Readability

In contrast to the rather large literature on L1 readability indexes, very little work has been done on readability indexes applied to second language (L2) students. A few such studies have investigated readability in languages other than English. For example, Nguyen and Henkin (1982) did so for Vietnamese, and Gilliam, Peña, and Mountain (1980) did so for Spanish. Moreover, Klare (1963) provided a survey of nine other early readability studies for French, German, Japanese, and Spanish.

For ESL, Haskell (1973) found that cloze procedure successfully distinguished among texts regardless of their length, the scoring method used, the deletion rate, and so forth. Hamsik (1984) found fairly strong associations (ranging from .78 to .82) between student performances on cloze tests and four different readability indexes. Brown (1998) showed that the mean performances of Japanese university students on the same 50 passages used in the present study correlated with the same readability estimates used in this study ranging from .48 to .55. Greenfield (1999) replicated Brown (1998) with different passages and found that the traditional L1 readability indexes correlated strongly with Japanese students' performances on cloze tests. Greenfield (2004) reported similar results and concluded that the traditional L1 readability indexes "are valid for EFL use" (p. 5).

Cloze Procedure and Readability

The first reference to cloze procedure was Taylor (1953), who studied the value of this sort of test for estimating the readability of reading materials used in U.S. public schools. Over the ensuing years, other key studies on cloze readability have included Bickley, Ellington, and Bickley (1970), Bormuth (1966, 1968), Brown (1998), Greenfield (1999, 2004), Miller and Coleman (1967), Moyle (1970), Ransom (1968), and Taylor (1957). All of these studies have shown that performances on cloze tests are at least somewhat related to readability.

Purpose

The primary purpose of this study was to explore the degrees of relationship between each of the L1 readability estimates and Russian EFL students' performances on actual cloze passages. To that end, the following research questions were posed:

1. Are randomly selected cloze tests reliable and valid tools for gathering data on passage difficulty?
2. To what degree are traditional first language readability indexes related to the average cloze scores for the same passages (when administered to Russian EFL students)?

METHOD

Participants

This study focused on the performances of 5170 Russian EFL students. The participants were selected as intact EFL classes from 38 different universities across Russia.¹ The participants ranged in

¹ We would like to thank all of our colleagues who helped at various stages of this project by administering the cloze tests at 38 universities in the following 25 towns and cities: Chelyabinsk, Kazan, Kolomna, Krasnodar, Moscow,

age from 14 to 45 with a mean of 18.59 (48 participants did not answer this question); 71.7% of the participants were female, 28.0% males, and 0.3% did not specify their gender. All 50 cloze procedures were administered in such a way that all students were randomly assigned across all testing sessions to particular cloze tests. The purpose of doing this was to ensure that the results of the different groups could reasonably be assumed to be comparable across the 50 passages. An average of 103.4 students took each cloze test (with individual passages ranging from 90 to 122 students).

The participants were mostly undergraduate students from non-linguistic universities and departments, though, some students were linguistics majors. Their levels of proficiency can generally be said to represent the English proficiency of university students in Russia who are studying subjects other than English or linguistics. While the participants in this study were not randomly sampled from all Russian university students, it can be said that the sample is fairly large and homogeneous with regard to the nationality, language background, and educational level of the students.

Materials

The 50 cloze procedures used in this study were developed by randomly selecting 50 books from the adult reading section of the Leon County Public Library in Tallahassee, Florida, and then randomly choosing a passage from each book to create a 350 to 450-word long passage, beginning from a semantically logical starting point. Clearly, these passages were not selected in a completely arbitrary manner, but they were selected so that they would form independent and cohesive passages. The resulting 50 cloze passages ended up ranging from 366 to 478 words with an average of 412.1 words in each passage. Based on random selection, the resulting set of 50 passages is assumed to represent the reading passages encountered in U.S. public library books.

To create the cloze passages, every 12th word was deleted from each text and was replaced with a standard length blank. A 12th word deletion pattern was chosen instead the more traditional 7th word deletion so that 30 items could be constructed far enough apart to minimize the effect of answering one item correctly (or incorrectly) on answering other items. In addition, one sentence was left unmodified by blanks at the beginning of each passage and one or two were left intact at the end of each passage. Additional spaces were then added at the top for the students' name, sex, age, native language, and country of passport. Directions were also given that explained what the students must do to fill in the blanks and how they would be scored. Sample directions and 12 sample items are shown in Appendix B, which was taken from Brown (1989).

An additional very short 10-item cloze passage was also created and attached to all 50 of the cloze tests. This 10-item cloze test was developed on the basis of pretesting reported in Brown (1989), using procedures similar to those applied in Brown (1988), so that only those blanks that had proven very effective in an item analysis were kept in the test. The purpose of this short anchor test was to provide a common metric for making comparisons among the fifty groups of students and for anchoring item response theory analyses, which will be reported in future studies.

Procedures

The data for this paper were collected by a large number of teachers at 38 universities in various locations throughout the Russian Federation (see footnote 5 for those locations). The cloze passages were randomly distributed in a manner that assured that all students had an equal chance of getting any of the 50 cloze tests. They were administered by the teachers in classroom. The directions were clarified as necessary, and a total of 25 minutes was allowed for completing both the 30-item and ten-item cloze tests. According to teacher feedback, 25 minutes was sufficient time for students to finish.

Exact-answer scoring was used in this research. Exact-answer scoring involves counting only the original word that had occupied the blank as a correct answer. We felt that this was justified because research has repeatedly shown high correlations between exact-answer scores and other more elaborate scoring procedures (Alderson, 1978, 1979; Brown, 1978, 1980).

Novocherkassk, Novosibirsk, Orenburg, Rostov/Don, Ryazan, Samara, Saransk, Saratov, Smolensk, Solykamsk, St. Petersburg, Surgut, Syktyvkar, Syzran, Taganrog, Togliatti, Tomsk, Ulyanovsk, Voronezh, and Yoshkar-Ola. For a list of the cooperating institutions, see Appendix A.

Analyses

The variables in this study were chosen because, they were known to be related to readability and because they were quantifiable. In other words, these variables were chosen because they might explain statistically the variations in readability levels of the cloze passages in this research. Only seven variables are included in this preliminary report: (a) six L1 readability indexes and (b) the means produced by the Russian EFL students who took these tests.

PRELIMINARY RESULTS

Table 2 shows the descriptive statistics for the 50 cloze tests ($k = 30$). These statistics include the mean, standard deviation (*SD*), minimum score obtained (Min), maximum score (Max), the number of participants who took the particular cloze (*N*), and the Cronbach alpha internal consistency reliability estimate for each test.

One obvious result which jumps out of Table 2 is that the means for the 50 cloze passages range from a low of 1.60 to a high of 12.82 out of 30. Since random selection promotes the equality of these groups in terms of overall English proficiency, the variation in means revealed in Table 2 is probably due to considerable variation in the difficulty levels of the passages involved. It is worth noting that these means are fairly low for tests with 30 items each. However, similarly low means have been commonly reported for cloze tests scored using the exact-answer method.

Notice also that the standard deviations range from a low of 2.27 to a high of 6.96. This range indicates substantial variation in the degree to which the students' scores were spread out around the means of these cloze passages. The minimum (Min) values are all 0. The maximum (Max) values ranged from 10 to 30, which indicates substantial variations in the ways these cloze passage scores spread out around their respective means. The number of participants on each cloze passage also ranged from 90 to 122.¹ The reliability of the 50 cloze tests likewise varied considerably. Notice that the lowest internal consistency Cronbach alpha reliability was .646, while the highest was .919. Such reliability estimates indicate the proportion of reliability or consistency in the scores. For example, .646 indicates that 64.6% of the variance in scores for that cloze test was reliable but also, by extension, that 35.4% was unreliable ($100\% - 64.6\% = 35.4\%$). Reliability estimates are important in any statistical research because a study can only be as reliable as the measures upon which it is based.

Table 2: *Descriptive Statistics for 50 Cloze Passages*

Test	Mean	SD	Min	Max	N	Reliability
1	6.78	3.99	0	16	120	0.745
2	7.06	4.94	0	19	102	0.853
3	3.94	3.71	0	14	103	0.811
4	9.82	6.12	0	21	105	0.887
5	6.54	4.38	0	22	106	0.822
6	5.34	4.19	0	16	102	0.828
7	8.07	6.22	0	20	103	0.896
8	3.13	3.67	0	24	101	0.859
9	4.08	3.67	0	23	105	0.808
10	3.77	4.24	0	22	102	0.869
11	5.74	4.53	0	17	101	0.845
12	9.27	4.86	0	20	115	0.834
13	3.30	3.89	0	17	105	0.855
14	5.10	4.70	0	17	107	0.866
15	8.10	5.60	0	21	106	0.893
16	2.30	2.70	0	11	115	0.773
17	2.55	2.29	0	10	109	0.646
18	1.60	2.27	0	15	100	0.775
19	6.15	5.08	0	30	102	0.882

¹ Note that, for reasons that we do not yet understand, Passage 1 originally had an unusually high number of participants with 170. We randomly selected 120 to remain in our data so that the sample sizes would all be approximately the same.

20	5.41	5.01	0	24	97	0.887
21	10.32	6.96	0	27	103	0.919
22	3.74	3.64	0	14	102	0.825
23	3.58	3.36	0	14	102	0.789
24	2.13	2.37	0	10	101	0.712
25	4.63	4.55	0	15	102	0.873
26	4.35	3.25	0	21	100	0.770
27	3.48	3.07	0	15	100	0.751
28	4.01	3.81	0	18	102	0.837
29	3.39	2.70	0	11	102	0.702
30	12.82	5.39	0	22	111	0.834
31	4.88	3.89	0	14	101	0.815
32	4.96	3.22	0	12	101	0.785
33	2.82	2.57	0	10	102	0.713
34	7.11	4.43	0	18	102	0.828
35	6.72	5.54	0	25	103	0.873
36	4.81	4.11	0	16	96	0.834
37	8.38	5.46	0	24	103	0.872
38	2.42	2.44	0	14	106	0.743
39	3.62	3.44	0	12	103	0.804
40	3.87	4.39	0	24	90	0.877
41	4.53	3.56	0	14	101	0.794
42	4.78	4.10	0	20	93	0.836
43	2.09	2.56	0	15	99	0.760
44	4.80	4.28	0	19	102	0.854
45	9.24	6.59	0	21	101	0.909
46	3.69	3.49	0	14	93	0.803
47	3.19	2.79	0	12	104	0.729
48	2.98	3.36	0	18	108	0.753
49	4.39	4.10	0	15	122	0.858
50	3.57	3.04	0	13	109	0.758

Table 3 displays the results for the L1 readability indexes examined in this research. They are arranged not by the passage numbers as they were in the previous table, but rather in order from the easiest to most difficult as indicated by the means in the last column. In other words, high means on passages (like the 12.82 for passage 30) indicate that the Russian students found them to be relatively easy, and low means (like the 1.60 for passage 18) indicate that the students found them to be relatively difficult.

Table 3: *L1 Readability Estimates and Russian Means for 50 Passages*

Passage	Flesch-			Gunning-			Mean
	Flesch	Kincaid	Fry	Gunning	Fog	Fog	
30	4.63	6.5	5	5.08	6.0	22	12.82
21	4.74	7.5	5	4.85	5.9	24	10.32
4	5.95	7.6	6	6.41	8.4	28	9.82
12	8.59	11.0	10	5.67	8.1	32	9.27
45	8.47	11.1	8	6.72	10.0	36	9.24
37	6.03	8.6	2	6.81	9.3	31	8.38
15	9.69	12.0	10	6.41	10.0	38	8.10
7	9.37	9.9	10	6.07	10.0	43	8.07
34	10.69	12.8	10	8.48	13.0	42	7.11
2	10.71	13.5	13	6.07	10.0	42	7.06
1	6.78	9.6	7	6.15	8.7	32	6.78
35	3.69	4.8	4	4.09	4.8	22	6.72
5	11.00	13.9	10	6.57	10.0	40	6.54

19	8.27	10.2	8	6.40	9.4	35	6.15
11	2.71	5.0	3	3.05	3.2	20	5.74
20	8.30	10.8	8	7.03	10.0	35	5.41
6	5.18	7.0	6	4.11	5.2	27	5.34
14	4.79	8.5	6	4.26	5.5	27	5.10
32	7.80	9.6	8	5.94	8.1	30	4.96
31	8.13	11.6	10	5.26	8.1	37	4.88
36	7.88	11.3	8	5.82	9.4	40	4.81
44	11.6	13.9	11	7.81	13.0	43	4.80
42	7.10	9.1	8	5.19	7.2	31	4.78
25	7.72	10.2	7	7.09	9.7	31	4.63
41	12.26	14.3	12	9.33	15.0	47	4.53
49	7.59	10.3	7	8.19	12.0	37	4.39
26	13.95	16.6	14	9.05	17.0	54	4.35
9	12.30	15.3	12	9.34	16.0	49	4.08
28	12.00	14.4	14	8.23	14.0	49	4.01
3	2.83	4.8	3	3.25	3.5	21	3.94
40	5.69	8.1	6	5.47	7.5	30	3.87
10	11.86	15.2	10	9.61	16.0	46	3.77
22	8.97	10.8	9	7.16	11.0	37	3.74
46	8.78	11.2	9	5.80	8.5	34	3.69
39	5.09	6.7	6	5.81	7.5	27	3.62
23	11.45	13.9	13	7.35	13.0	46	3.58
50	18.51	21.3	15	13.48	25.0	64	3.57
27	9.36	10.0	9	7.20	11.0	38	3.48
29	13.58	16.0	11	11.00	17.0	46	3.39
13	10.65	12.1	10	8.83	14.0	40	3.30
47	9.99	11.9	9	8.24	13.0	40	3.19
8	8.46	11.2	8	7.83	11.0	36	3.13
48	8.51	11.2	8	6.95	12.0	44	2.98
33	13.82	16.3	12	11.01	21.0	59	2.82
17	15.60	20.4	14	9.78	19.0	58	2.55
38	11.01	12.9	11	8.13	13.0	42	2.42
16	8.90	13.0	9	8.99	16.0	50	2.30
24	10.69	13.1	10	8.95	14.0	40	2.13
43	11.51	13.9	10	9.72	15.0	43	2.09
18	9.69	12.7	12	6.06	9.7	40	1.60

The remaining columns in Table 3 show the readability estimates for each passage using the Flesch, Flesch-Kincaid, Fry, Gunning, Fog, and Gunning-Fog indexes. All of these indexes, except for the Gunning-Fog index, are meant to be on scales that represent grade levels in U.S. schools. It is interesting that they are fairly comparable in some cases, thus indicating similar relative difficulties for the passages. In other words, passages that appear to be relatively easy on one index also tend to be relatively easy on the other ones as well, while passages that appear to be relatively difficult on one index are also relatively difficult on the others.

Table 4 shows the correlation coefficients for all possible pairs of the L1 readability estimates in this study and the Russian means. The darker cells that spread across the table diagonally indicate the correlations of each variable with itself, which in each case is of course 1.00. Note then that while a perfect correlation is 1.00 and total lack of correlation would equal .00, all the values in the table are somewhere between the two extremes. Notice also that the correlations in the last light gray column are negative values indicating that the means range from easy to difficult in the opposite direction from the L1 readability indexes. In other words, cloze passages with high means are relatively easy while those with low means are difficult. This is the opposite (thus negative) from the readability indexes where high values indicate difficult passages (suitable for higher grades) and low values indicate easy

passages (suitable for lower grades). It turns out that the L1 indexes are all moderately to highly correlated with each other (between .70 and .98) depending on which one is examined. It also turns out that there is a relatively low degree of relationship between the Russian means and the various L1 readability indexes. All in all, these L1 indexes are more highly related to each other than they are to the Russian means.

Table 4: *Correlation Coefficients L1 Readability Indexes and Russian Mean Difficulty¹*

	Flesch	Flesch-Kincaid	Fry	Gunning	Fog	Gunning-Fog	Means
Flesch	1.00	.98	.92	.89	.93	.95	-.42
Flesch-Kincaid		1.00	.90	.87	.92	.95	-.44
Fry			1.00	.70	.78	.88	-.41
Gunning				1.00	.98	.87	-.45
Fog					1.00	.95	-.48
Gunning-Fog						1.00	-.48
Russian Means							1.00

DISCUSSION

The discussion will now return to the original research questions and will address each question separately. The implications of these findings for Russian EFL readability estimations will be explored in the Conclusions section.

1. Are Randomly Selected Cloze Tests Reliable and Valid Tools for Gathering Data on Passage Difficulty?

In terms of reliability, Table 2 indicates that the cloze passages in this study are reasonably reliable, though individual tests varied somewhat with Cronbach alpha reliability estimates ranging from a moderate .646 to a relatively high .919. This means that these passages ranged from being about two-thirds reliable (64.6%) to being more than nine-tenths reliable (91.9%). To some degree, such variation in reliability can be related to the distributions of scores. The magnitudes of the means (some of which were as low as 1.60) and standard deviations (many of which were almost as large as their corresponding means) indicate that many of these distributions were probably positively skewed—a fact that would tend to depress the values of Cronbach alpha. Nonetheless, these estimates represent the reliability of these cloze tests when used under these conditions with these students.

In terms of validity, an argument can be built for the validity of the scores on these 50 cloze tests as follows. Given that the cloze passages were constructed from randomly selected public library books and that the items for the passages were semi-randomly selected (i.e., every 12th word deletion), sampling theory would indicate that the passages form a representative sample of the English language found in those library books. Therefore, it can be argued that the items form a representative sample of the blanks that can be created from public library books. Given that the validity of the scores from a set of items is defined as the degree to which they are measuring what they purport to measure, the validity argument here is that these cloze items have a high degree of content validity because they can be said to form representative samples of the universe of all possible items (after Cronbach, 1970), if that universe is defined as blanks created from the written language found in an American public library. [For much more on the reliability and validity of these passages, see Brown, 1993; for more on test reliability and validity issues, see Brown, 2005.]

¹ Note that all of the correlation coefficients in Table 4 are statistically significant at $p < .01$ (one-tailed).

2. To What Degree Are Traditional First Language Readability Indexes Related to the Average Cloze Scores for the Same Passages (When Administered to Russian EFL Students)?

Tables 3 and 4 both show some degree of relationship between each of the L1 readability indexes and the Russian means. Table 3 allows readers to actually inspect these relationships. However, Table 4 shows the degree to which the L1 readability indexes are moderately to very highly correlated with each other (ranging from .70 to .98). Thus, the L1 readability indexes appear to be moderately to highly interrelated, which makes sense given that they are all based on the same sorts of counts of syllables, words, sentences, etc. Table 4 also shows that L1 readability indexes are somewhat related to the performances of Russian EFL students as indicated by the correlations of -.41 to -.48. Remember that these coefficients are on a scale from no correlation (.00) to perfect correlation (1.00) and that they are negative values because the L1 readability and Russian mean scales indicate passage difficulty in opposite directions. Naturally, all of this can be said to be true only for Russian university EFL students as sampled in this study and for the cloze passages used here.

FUTURE DIRECTIONS

The results of this preliminary study indicate that there is certainly reason to move forward with further analyses. It has shown that a variety of L1 readability indexes for this set of 50 passages are somewhat correlated with the average performances of Russian university students on cloze versions of those same passages. This finding is important to think about. Why are the L1 readability indexes only somewhat related to the Russian EFL means? This lack of relationship could be due to any of the following: (a) that these L1 readability estimates are fine indicators of passage readability for native speakers but not for Russian EFL learners; (b) that the cloze passages are measuring something different from the simple readability measured by the L1 indexes; (c) that the Russian EFL learner's scores on the cloze passages are measuring something much more complex than simple readability—something like the students' *overall proficiency* levels rather than the *reading* difficulty of the passages.

We are hoping that the analyses that we are planning to perform on these data in the future will help us to better understand these and other aspects of these cloze passages. In particular, the directions we anticipate pursuing at the moment should lead us to answers to the following research questions:

1. What other linguistic text variables (e.g., word length, word frequency for each blank, the length of the sentence in which the blank is found, whether the word is of Germanic or Latinate origin, etc.) should be included in such research?
2. How well do those linguistic text variables predict Russian EFL performances at the passage level? At the item level? And in what combinations? [Using item-response theory, factor analysis, multiple-regression analyses, and structural equation modeling—all of which will be explained in subsequent papers.]
3. What hierarchies of difficulty are found at the passage level for any of the linguistic variables (separately or combined) that would have implications for second language acquisition research? Similarly, what hierarchies of difficulty are found at the item level?
4. What differences and similarities would occur if the results of this study were compared with the similar data gathered in Japan? With students from other language groups? With students at other levels of study? Or at other ages?

REFERENCES

- Alderson, J. C. (1978). *A study of the cloze procedure with native and non-native speakers of English*. Unpublished doctoral dissertation, University of Edinburgh.
- Alderson, J. C. (1979). Scoring procedures for use on cloze tests. In C. A. Yorio, K. Perkins, & J. Schachter (Eds.) *On TESOL '79* (pp. 193-205). Washington, DC: TESOL.
- Bickley, A. C., Ellington, B. J., & Bickley, R. T. (1970). The cloze procedure: A conspectus. *Journal of Reading Behavior*, 2, 232-234.
- Bormuth, J. R. (1966). Readability: A new approach. *Reading Research Quarterly*, 1, 79-132.

- Bormuth, J. R. (1967). Comparable cloze and multiple-choice comprehension tests scores. *Journal of Reading, 10*, 291-299.
- Brown, J. D. (1978). *Correlational study of four methods for scoring cloze tests*. Unpublished MA thesis, University of California, Los Angeles.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *Modern Language Journal, 64*, 311-317.
- Brown, J. D. (1988). Tailored cloze: Improved with classical item analysis techniques. *Language Testing, 5*, 19-31.
- Brown, J. D. (1989). Cloze item difficulty. *JALT Journal, 11*, 46-67.
- Brown, J. D. (1998). An EFL readability index. *JALT Journal, 20*(2), 7-36.
- Brown, J. D., Chen, Y., & Wang, Y. (1984). An evaluation of native-speaker self-access reading materials in an EFL setting. *RELC Journal, 15*, 75-84.
- Carrell, P. L. (1987). Readability in ESL. *Reading in a Foreign Language, 4*, 21-40.
- Chall, J. S. (1958). *Readability: An appraisal of research and application*. Columbus, Ohio: Ohio State University.
- Cronbach, L. J. (1970). *Essentials of psychological testing* (pp. 145-146). New York: Harper & Row.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221-233.
- Fry, E. B. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading, 21*, 242-252.
- Fry, E. B. (1985). *The NEW reading teacher's book of lists*. Englewood Cliffs, NJ: Prentice-Hall.
- Gilliam, B., Peña, S. C., & Mountain, L. (1980). The Fry graph applied to Spanish readability. *The Reading Teacher, 33*, 426-430.
- Greenfield, G. (1999). *Classic readability formulas in an EFL context: Are they valid for Japanese speakers?* Ed.D. dissertation, Temple University.
- Greenfield, J. (2004). Readability formulas for EFL. *JALT Journal, 26*(1), 5-24.
- Hamsik, M. J. (1984). *Reading, readability, and the ESL reader*. Unpublished doctoral dissertation, Florida State University, Tallahassee, FL.
- Haskell, J. (1973). *Refining the cloze testing and scoring procedures for use with ESL students*. Unpublished doctoral dissertation, Columbia University, New York.
- Klare, G. R. (1963). *The measurement of readability*. Ames, IA: Iowa State University.
- Klare, G. R. (1984). Readability. In R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 681-744). New York: Longman.
- Larson, R. (1987). How Readability was created. In Scandinavian PC Systems. *Readability program for the IBM PC, XT and AT* (pp. 8-1 to 8-20). Rockville, MD: Scandinavian PC Systems.
- Miller, G. R., & Coleman, E. B. (1967). A set of thirty-six prose passages calibrated for complexity. *Journal of Verbal Learning and Verbal Behavior, 6*, 851-854.
- Moyle, D. (1970). Readability—the use of cloze procedure. In J. Merrit (Ed.), *Reading and the curriculum* (pp. 159-168). London: Ward Lock Educational.
- Nguyen, L. T., & Henkin, A. B. (1982). A readability formula for Vietnamese. *Journal of Reading, 26*, 243-251.
- Ransom, P. E. (1968). Determining reading levels of elementary school children by cloze testing. In J. A. Figurel (Ed.), *Forging ahead in reading* (pp. 477-482). Newark, DE: International Reading Association.
- Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly, 30*, 414-438.
- Taylor, W. L. (1957). "Cloze" readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology, 41*, 19-26.
- Zakaluk, B. L., & Samuels, S. J. (1988). *Readability: Its past, present, and future*. Newark, DE: International Reading Association.
- Zamanian, M., & Heydari, P. (2012). Readability of texts: State of the art. *Theory and Practice in Language Studies, 2*(1), 43-53.

APPENDIX A: PARTICIPATING UNIVERSITIES

We owe a great debt of gratitude to our colleagues at the following 38 institutions who helped us gather the data upon which this study is based. Without their cooperation and help, and the efforts of their students, this research project would simply not exist.

1. Chelyabinsk Law Institute
2. Chelyabinsk State University
3. International Market Institute (Samara)
4. Kazan branch of the Russian International Academy for Tourism
5. Kazan Military Institute
6. Kazan State Technical University
7. Kolomna State Pedagogical University
8. Krasnodar State University
9. Mordovian State University
10. Lomonosov Moscow State University
11. Novocherkassk Polytechnic Institute
12. Novosibirsk State University
13. Orenburg State University
14. Presidential Cadet College (Orenburg)
15. Rostov/Don Institute of Management, Business and Law
16. Rostov/Don State University
17. Ryazan State University
18. Korolyov Samara Aerospace university
19. Samara State Academy of Social Sciences and Humanities
20. Samara State University
21. Samara State University of Architecture and Civil Engineering
22. Saratov State Pedagogical University
23. Saratov State University
24. Smolensk University for the Humanities
25. Solykamsk State Pedagogical University
26. South-Ural State University
27. St. Petersburg State University
28. Surgut State University
29. Syktyvkar State University
30. Syzran branch of Samara State Technical University
31. Taganrog Institute of Management and Economics
32. Taganrog State Pedagogical University
33. Togliatti Academy of Management
34. Tomsk Polytechnic University
35. Ulyanovsk State University Institute for International Relations
36. Volga State University of Technology (former Mari State Technical University)
37. Voronezh State University
38. Voronezh State University of Architecture and Civil Engineering

APPENDIX B: EXAMPLE CLOZE PASSAGE

(ADAPTED FROM BROWN, 1989)

Name _____ Native Language _____
(Last) (First)
Sex _____ Age _____ Country of Passport _____

DIRECTIONS:

1. Read the passage quickly to get the general meaning.
2. Write *only one* word in each blank. Contractions (example: *don't*) and possessives (*John's bicycle*) are one word.
3. Check your answers.

NOTE: Spelling will *not* count against you as long as the scorer can read the word.

EXAMPLE: The boy walked up the street. He stepped on a piece of ice. He fell
(1) _____, but he didn't hurt himself.

A FATHER AND SON

Michael Beal was just out of the service. His father had helped him get his job at Western. The (1) _____ few weeks Mike and his father had lunch together almost every (2) _____. Mike talked a lot about his father. He was worried about (3) _____ hard he was working, holding down two jobs.

"You know," Mike (4) _____, "before I went in the service my father could do just (5) _____ anything. But he's really kind of tired these days. Working two (6) _____ takes a lot out of him. He doesn't have as much (7) _____. I tell him that he should stop the second job, but (8) _____ won't listen.

During a smoking break, Mike introduced me to his (9) _____. Bill mentioned that he had four children. I casually remarked that (10) _____ hoped the others were better than Mike. He took my joking (11) _____ and, putting his arm on Mike's shoulder, he said, "I'll be (12) _____ if they turn out as well as Mike."
(continues ...)