Coll. Antropol. 39 (2015) 1: 11–20
Original scientific paper

Development of SMOG-Cro Readability Formula for Healthcare Communication and Patient Education

Sanja Brangan

University of Zagreb, School of Medicine, »Andrija Štampar« School of Public Health, Department of Educational Technology, Zagreb, Croatia

ABSTRACT

Effective communication shows a positive impact on patient satisfaction, compliance and medical outcomes, at the same time reducing the healthcare costs. Written information for patients needs to correspond to health literacy levels of the intended audiences. Readability formulas correlate well with the reading and comprehension tests but are considered an easier and quicker method to estimate a text difficulty. SMOG readability formula designed for English language needs to be modified if used for texts in other languages. The aim of this study was to develop a readability formula based on SMOG, that could be used to estimate text difficulty of written materials for patients in Croatian language. Contrastive analysis of English and Croatian language covering a corpus of almost 100,000 running words showed clear linguistic differences in the number of polysyllabic words. The new formula, named SMOG-Cro, is presented as an equation: SMOG-Cro =2+$\sqrt{4^{+}}$ syllables, with the score showing the number of years of education a person needs to be able to understand a piece of writing. The presented methodology could help in the development of readability formulas for other languages. We hope the results of this study are soon put into practice for more effective healthcare communication and patient education, and for development of a health literacy assessment tool in Croatian language.

Key words: *doctor-patient communication, patient education, readability formulas, SMOG, health literacy, patient-centred health care*

Introduction

Communication is vital to doctor-patient relationship, and its quality may greatly affect this relationship at the individual level, at the community level, but also at the level of the entire healthcare system. Effective communication shows a positive impact on patient satisfaction, compliance and medical outcomes, at the same time reducing the healthcare costs[1–4]. In other words, the overall quality of the healthcare services may depend on the quality of communication between patients and healthcare providers.

It is important to achieve adequate communication during doctor-patient consultation, for diagnostic and therapeutic purposes, for informed consent, for clinical trials procedure, for patient education and counseling, and similar. When communicating with the general population about health issues, very often experts with various professional backgrounds are involved including, apart from medical doctors, psychologists, sociologists, social workers, educators, and journalists, just to name a few. Whatever the context and the specific health topic, the two or more parties involved would have a mutual aim of understanding each other in the first place. Only then would the patients, or the general population, become competent enough to make valid decisions related to their health. This process of empowering the patients is in fact at the core of the patient-centred health care, which values patients as the key participants in the process of shared decision-making, respecting their needs and preferences, and trying to look at the problem from their perspective, as opposed to the traditional »paternalistic« approach, where the conflict between medical authority and patient autonomy is seen as fundamental to the doctor-patient relationship[5].

However, data from the healthcare system show a disturbing fact that over 50% of patients do not understand their doctors[6]. Furthermore, over 300 studies have shown that written information for patients is too difficult to be understood by the intended audience[7]. The text difficulty of those written materials for patients simply does not match the literacy levels of the audience. In the healthcare

Received for publication July 21, 2013

context, a concept of health literacy has been developed, but due to its complexity, various definitions have been published in the literature. The 1999 definition stated that health literacy is »the constellation of skills, including the ability to perform basic reading and numerical tasks required to function in the health care environment«[8], while the often cited one from 2004 defines health literacy as »the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions«[7]. A simpler, and more goal-oriented definition claims that »health literacy is a shared responsibility in which patients and providers each must communicate in ways the other can understand«[9].

Written information for patients needs to be well understood by patients so they can make the appropriate health decisions. Many authors on the subject recommend involving patients at an early stage of written information development, and mandatory checking of the final version with the intended audience so that the text difficulty matches the literacy levels of users. To check the literacy levels of individual patients, some reading tests and tests of functional health literacy have been developed. Reading tests, such as WRAT (Wide Range Achievement Test) and REALM (Rapid Estimate of Adult Literacy in Medicine) are based on isolated words of increasing difficulty and would indentify patients with limited reading skills. These tests have been designed for English language and cannot be readily used for speakers of other languages even when translated[7]. A frequently used test of functional health literacy in adults, TOFHLA[10] and its shorter version S-TOFHLA[11], have been designed for English and Spanish language, and the translations have been later provided for other languages, such as Hebrew[12], Brazilian Portuguese[13], Serbian[14,15], Chinese[16], Mandarin[17], Korean[18], German, Italian, and French[19] but with necessary cultural adaptations and not clear indication of the text difficulty levels of the translations.

A different type of test may be applied to estimate the text difficulty, which would take into account the literacy levels of the readers, and at the same time providing results that correlate well with the reading tests. These »tests« are readability formulas, which are actually mathematical equations that provide a rough estimate of text difficulty but are a very cheap, quick, easy-to-use, and an efficient method for measuring text readability. Around 200 readability formulas have been developed over the past nine decades for different languages, based mostly on word and sentence length but some requiring frequency lists, and they have stood the test of time[20]. Readability formulas applied for health information materials, as mentioned in the literature, are SMOG readability formula, Flesch Reading Ease, Flesch-Kincaid Grade Level, Gunning's Fog Index, Fry Readability Scale, etc.[6,7,9,20]. The word »readability« itself has been defined by those researching and designing readability formulas as »the ease of understanding or comprehension due to the style of writing«[21] or »the ease of reading words and sentences«[22], and is seen as an attribute of clarity.

SMOG readability formula has been used extensively for analysis of health written materials designed for patients. It is a manual method of readability testing, introduced by McLaughlin in 1969, which estimates the years of education needed to understand a piece of writing, and predicts 90–100% comprehension[23]. It is a quick, consistent and easy-to-use method, based on the number of polysyllabic words in a text. The name of the formula could be associated either with »smog«, as its author indicated referring to British weather and his birthplace London, or with »Simple Measure of Gobbledygook«, from which the acronym SMOG has been created[24]. Furthermore, the author himself called it »a complimentary allusion to Gunning's Fog Index«, a readability formula developed in 1952[25].

This paper presents the development of SMOG-Cro, a readability formula based on McLaughlin's SMOG formula, that could be used for estimating difficulty of the written materials designed for patients in Croatian language in the context of healthcare communication and patient education. An in-depth analysis of the differences between English and Croatian language in the parameters needed for the formula, performed on a corpus of almost 100,000 running words, provides also a model for development of readability formulas in other languages.

Materials and Methods

A contrastive analysis of English and Croatian language was done on texts with expected varying difficulty. Ninety samples of 30 sentences were taken from the following four sets: 33 samples from books of fiction by popular authors; 24 samples of feature articles from SETimes online journal[26]; 18 samples from popular science books; and 15 samples from research articles published in the Journal of the American Medical Association (JAMA). The analysis covered each sample published in English language and its translation into Croatian language published since 1995, providing a corpus of 2,700 sentences in each language or a total of almost 100,000 running words.

As indicated by the studies on readability, the samples were taken from the beginning, the middle, and the end of the books and research papers, leaving out the first and the last sentence, except for feature articles, where the samples were much shorter and just the first or the last sentence was omitted alternately. The titles or subtitles were omitted as well, and abbreviations and numbers spelled out for proper calculation of the number of syllables. Sentences that were not found in 1:1 ratio in their translation were skipped, enabling analysis of 30 sentences in each sample.

Total number of words per sample was calculated using 'wordcount' option from Word for Windows XP. Total numbers of syllables and words with three or more syllables per sample were calculated using freely available online readability calculator[27], which was shown to be the most reliable when compared to manual calculation on a random sample by native speakers of English and Croatian

language. Total number of words of four or more syllables for samples in Croatian language was calculated manually by a native speaker of Croatian language. SMOG readability formula was calculated using Excel for Windows XP, and the median values and ranges of scores in Croatian language corresponding to each score obtained for English language were calculated by SPSS statistical package for data analysis.

The following equation was used as the SMOG readability formula:

$$\text{SMOG} = 3 + \sqrt{3^{+}\ \text{syllables}}$$

presented here from a descriptive wording given by McLaughlin »3+ square root of polysyllable count«, or from his more precise but lengthier description, which could be summarized as: »add 3 to the approximate square root of the number of polysyllabic words counted«[23], where the definition of »a polysyllabic word« is »a word of at least three syllables«. Samples in Croatian language were additionally calculated for the number of words with at least four syllables. The obtained SMOG scores indicate the number of years of education needed to understand a text.

Results

The corpus analyzed in this study covered almost 100,000 running words consisting of 90 samples of 30 sentences or 51,160 running words in English language, and 90 samples of 30 sentences or 46,533 running words in Croatian language, translations published since 1995. The corpus consisted of four sets of texts with varying expected difficulty. Set 1 (samples 1–33) included 33 samples from books of fiction by popular authors, such as J.K. Rowling, Robin Cook, Dan Brown, Stephen King, Michael Crichton, Dean Koontz, David Lodge, and John Grisham (Table 1). The number of syllables and words with at least three syllables was consistently higher in samples in Croatian language. The percentage of words with at least three syllables was 8.6% and 30.2% for English and Croatian language, respectively. SMOG scores were also consistently higher for samples in Croatian language, with the lowest scores found in books by Dan Brown. Set 2 (samples 34–57) included 24 samples from feature articles of SETimes online journal, which publishes news and views of Southeast Europe in several languages (Table 2). The number of syllables and words with at least three syllables was again consistently higher in samples in Croatian language. The percentage of words with at least three syllables was 19.9% and 40.6% for English and Croatian language, respectively. SMOG scores were also consistently higher for samples in Croatian language. Set 3 (samples 58–75) consisted of 18 samples from popular science books by authors Bryan Sykes, Richard Dawkins, Bill Bryson, Steve Jones, Misha Glenny, and Stephen Hawking (Table 3). Again, the number of syllables and polysyllabic words, and the SMOG scores were consistently higher for samples in Croatian language. The percentage of words with at least three syllables was 14.6% and 38% for English and Croatian language, respectively. Set 4 (samples 76–90) consisted of 15 samples from JAMA original research papers with topics on inguinal hernia, hyperlipidemia, atrial fibrillation, sinusitis, and migraine, published between 2000 and 2006, and the samples in Croatian were taken from the published translations found in the Croatian edition of JAMA. Again, the number of syllables and polysyllabic words, and the SMOG scores were consistently higher for samples in Croatian language (Table 4). The percentage of words with at least three syllables was 23.3% and 49% for English and Croatian language, respectively.

As shown by Tables 1–4, the number of polysyllabic words in Croatian language was much closer to that found for English language when the term 'polysyllabic' is redefined for Croatian language to mean 'words with at least four syllables'. When calculated for totals in all 90 samples, the number of words with at least three syllables was 8,574 and 18,386 for English and Croatian language, respectively, and the number of words with at least four syllables was 9,728 for Croatian language. The SMOG scores were consistenly higher for samples in Croatian language in all sets when the original formula for English language was applied. Table 5 shows differences in SMOG scores for Croatian language when the two definitions of a polysyllabic word are used – one taking a count of words with at least three syllables, and the other taking a count of words with at least four syllables. Mean and median values are given for scores in Croatian language for each SMOG score found in English language (Table 5).

Figure 1 shows the SMOG scores obtained in Croatian language for each SMOG score found in English language in more than five samples, indicating a clear difference of 1 (Figure 1). This has led to the modification of the original SMOG formula designed for English language in the form of $3+\sqrt{3^{+}}$ syllables to the formula for Croatian language in the form of $2+\sqrt{4^{+}}$ syllables, which we named SMOG-Cro, with the meaning of Croatian SMOG readability formula, where a coefficient of 2 is added to the nearest square root of the total number of words with at least four syllables. The SMOG-Cro readability formula is to be applied for texts of 30 sentences, indicating a rough estimate of years of education a person needs to understand a piece of writing.

Discussion

Contrastive analysis of the studied samples showed the differences between English and Croatian language in the number of words, number of syllables and polysyllabic words, and consequently, in the SMOG scores calculated using the original SMOG readability formula designed for English language (Tables 1–4). The total number of words in all 90 samples was 51,160 running words compared to 46,533 in English and Croatian, respectively, while the total number of syllables was 84,772 in English and 105,689 in Croatian language. That gives the average word length expressed in the number of syllables 1.6 for English, and 2.2. for Croatian language. The total number of words with at least three syllables was 8,574 and 18,386 for English and Croatian language, respectively, and the

TABLE 1
TOTAL NUMBER OF WORDS, SYLLABLES, WORDS WITH 3+ AND 4+ SYLLABLES, AND SMOG SCORES FOR ENGLISH AND CROATIAN LANGUAGE IN 33 SAMPLES OF SET 1

Sample number	Total words		Total syllables		3+ syll.		4+ syll.	SMOG score	
	EN	CRO	EN	CRO	EN	CRO	CRO	EN	CRO
1	487	460	673	954	25	133	66	8	15
2	401	369	540	696	22	87	38	8	12
3	333	324	448	634	22	89	40	8	12
4	451	389	677	814	55	132	60	10	14
5	279	233	371	435	20	53	29	7	10
6	259	219	368	459	23	71	23	8	11
7	416	373	638	796	50	130	55	10	14
8	494	444	748	927	66	140	78	11	15
9	323	288	503	635	54	111	52	10	14
10	466	390	748	871	64	150	79	11	15
11	299	255	399	510	23	83	27	8	12
12	252	216	349	423	23	64	24	8	11
13	200	191	276	383	13	58	18	7	11
14	156	140	225	271	12	36	12	6	9
15	249	244	371	494	30	75	36	8	12
16	183	164	235	330	8	48	29	6	10
17	220	220	281	439	8	65	28	6	11
18	163	171	220	311	8	40	15	6	9
19	590	481	776	906	28	118	50	8	14
20	329	270	445	516	24	72	27	8	11
21	404	325	522	627	17	82	31	7	12
22	272	240	367	482	22	72	28	8	11
23	308	285	406	560	23	92	44	8	13
24	278	250	360	499	13	71	36	7	11
25	442	395	640	789	39	117	48	9	14
26	581	551	821	1118	58	190	65	11	17
27	360	344	452	635	12	87	22	6	12
28	602	564	928	1256	72	216	104	11	18
29	632	559	895	1094	56	150	56	10	15
30	625	553	937	1183	81	199	90	12	17
31	356	380	498	748	26	110	46	8	13
32	265	245	324	426	12	55	15	6	10
33	174	168	228	313	12	37	11	6	9

3+ syll. – words with at least three syllables, 4+ syll. – words with at least four syllables
EN – English language, CRO – Croatian language

total number of words with at least four syllables was 9,728 for Croatian language. The average number of characters per syllable was 3 for English and 2.4 for Croatian language (data not presented), which further showed a specific structure of words in the two languages – on the average, the words in Croatian language are longer but with shorter syllables.

The observed different word length in the two languages affected the calculated SMOG scores, which were much higher for Croatian language in all studied samples. Both mean and median SMOG values for Croatian language were higher than for English language, but comparable when the number of words with at least four syllables was included in the formula (Table 5). This has led to a definition of »a polysyllabic word« in Croatian language, for the purposes of SMOG formula calculation, as 'a word of at least four syllables'. When a coefficient 3 was then added to the nearest square root of the number of thus defined

TABLE 2
TOTAL NUMBER OF WORDS, SYLLABLES, WORDS WITH 3+ AND 4+ SYLLABLES, AND SMOG SCORES FOR ENGLISH AND CROATIAN LANGUAGE IN 24 SAMPLES OF SET 2

Sample number	Total words		Total syllables		3+ syll.		4+ syll.	SMOG	
	EN	CRO	EN	CRO	EN	CRO	CRO	EN	CRO
34	530	457	840	965	85	152	71	12	15
35	513	489	815	1093	73	180	97	12	16
36	716	594	1246	1441	137	270	136	15	19
37	528	444	720	875	33	119	50	9	14
38	628	529	977	1146	82	182	79	12	16
39	659	597	1141	1358	135	239	132	15	18
40	859	810	1550	1839	193	327	185	17	21
41	620	547	1075	1324	126	239	121	14	18
42	605	560	1096	1272	130	220	121	14	18
43	538	483	937	1074	99	176	86	13	16
44	714	643	1342	1551	179	286	171	16	20
45	710	649	1283	1496	161	262	162	16	19
46	716	638	1200	1479	129	263	147	14	19
47	600	506	1010	1183	115	222	110	14	18
48	740	683	1355	1602	164	295	174	16	20
49	538	486	897	1144	100	208	107	13	17
50	729	642	1344	1495	183	254	153	17	19
51	717	650	1258	1541	154	274	159	15	20
52	624	571	1136	1330	147	232	121	15	18
53	911	862	1492	1926	156	335	159	15	21
54	666	555	1162	1331	129	241	125	14	19
55	660	596	1159	1422	122	262	152	14	19
56	871	817	1635	1960	212	372	190	18	22
57	797	680	1428	1568	180	273	151	16	20

3+ syll. – words with at least three syllables, 4+ syll. – words with at least four syllables
EN – English language, CRO – Croatian language

polysyllabic words, a difference was still observed in SMOG scores for Croatian language when compared to English language. The selected seven SMOG scores (SMOG scores 6, 8, 11, 12, 14, 15, and 16) found for English language in more than five samples, when compared to SMOG scores obtained for Croatian language, showed a clear difference of 1 (Figure 1), which was included in the calculation and showed that the SMOG readability formula for English language, in the form of SMOG $=3+\sqrt{3^+}$ syllables, corresponds to the SMOG readability formula in Croatian language, which we named SMOG-Cro, in the form of SMOG-Cro $=2+\sqrt{4^+}$ syllables. Since the SMOG formula has been primarily designed for manual calculation, without the need of a computer program to calculate the number of polysyllabic words, the ease of manual calculation present for texts in English is kept this way for texts in Croatian language as well.

A similar approach, with varying methodology, has been applied in the studies modifying the original SMOG formula for languages other than English. A literature review showed that SMOG formula has been modified for the following languages: Spanish, French[28]; Turkish[29]; and Greek[30]. Contreras et al.[28] called their formulas for Spanish and French language »SOL formulas«, explaining that »sol« means »sun« in English, and that »it was time for the SOL to come out«. They analyzed 264 samples of 30 sentences from written materials of different reading difficulty available in English, Spanish, and French, and observed that SMOG scores are systematically higher in Spanish than in French, and higher in French than in English. To explain the differences, they calculated the number of polysyllabic words and the average sentence length but only on a random sample of 90 sentences of one type of writing, namely The Little Prince, which was »literally« translated. One cannot but argue that any conclusions drawn from such an approach would be biased and not really representing the composition of the three languages in general. For instance, they found that the percentage of polysyllabic words, i.e. words with at least three syllables, is 8.08% in English language, with Spanish almost three-fold the number in English, and two-fold the

TABLE 3
TOTAL NUMBER OF WORDS, SYLLABLES, WORDS WITH 3+ AND 4+ SYLLABLES, AND SMOG SCORES FOR ENGLISH AND CROATIAN LANGUAGE IN 18 SAMPLES OF SET 3

Sample number	Total words		Total syllables		3+ syll.		4+ syll.	SMOG	
	EN	CRO	EN	CRO	EN	CRO	CRO	EN	CRO
58	736	683	1122	1563	79	285	133	12	20
59	649	541	1005	1222	89	209	116	12	17
60	600	529	983	1204	100	213	108	13	18
61	621	546	977	1170	89	182	87	12	16
62	634	572	1018	1214	100	200	99	13	17
63	625	522	1087	1222	124	222	123	14	18
64	519	426	762	881	52	137	54	10	15
65	626	557	986	1185	77	195	93	12	17
66	600	528	876	1100	63	181	78	11	16
67	677	600	972	1275	62	216	82	11	18
68	576	492	891	1079	70	181	85	11	16
69	723	630	1109	1385	91	241	121	13	19
70	807	719	1353	1640	153	293	139	15	20
71	791	713	1340	1695	149	313	191	15	21
72	865	827	1506	1982	169	351	215	16	22
73	657	557	947	1142	72	173	69	11	16
74	834	732	1287	1589	121	255	143	14	19
75	656	632	1143	1434	125	256	132	14	19

3+ syll. – words with at least three syllables, 4+ syll. – words with at least four syllables
EN – English language, CRO – Croatian language

TABLE 4
TOTAL NUMBER OF WORDS, SYLLABLES, WORDS WITH 3+ AND 4+ SYLLABLES, AND SMOG SCORES FOR ENGLISH AND CROATIAN LANGUAGE IN 15 SAMPLES OF SET 4

Sample number	Total words		Total syllables		3+ syll.		4+ syll.	SMOG	
	EN	CRO	EN	CRO	EN	CRO	CRO	EN	CRO
76	683	717	1325	1805	180	347	189	16	22
77	715	683	1331	1681	164	326	177	16	21
78	762	698	1384	1701	175	322	182	16	21
79	681	759	1309	1803	160	340	206	16	21
80	680	714	1271	1653	155	298	175	15	20
81	924	941	1751	2223	221	421	224	18	24
82	678	686	1396	1854	185	352	220	17	22
83	796	801	1405	2027	147	400	207	15	23
84	650	643	1308	1722	177	338	222	16	21
85	769	642	1530	1713	212	344	223	18	22
86	715	639	1343	1677	153	340	188	15	21
87	864	727	1658	1934	215	394	238	18	23
88	634	597	1182	1604	143	310	211	15	21
89	764	716	1345	1805	134	330	233	15	21
90	611	576	1103	1556	123	305	224	14	20

3+ syll. – words with at least three syllables, 4+ syll. – words with at least four syllables
EN – English language, CRO – Croatian language

TABLE 5
RANGE, MEAN AND MEDIAN VALUES FOR SMOG SCORES IN CROATIAN LANGUAGE FOR WORDS WITH 3+ AND 4+ SYLLABLES, BY EACH SMOG SCORE FOUND IN SAMPLES IN ENGLISH LANGUAGE

EN		CRO					
SMOG score	N	3+ syllables			4+ syllables		
		SMOG range	Mean SMOG score	Median SMOG score	Range	Mean SMOG score	Median SMOG score
6	7	9–12	10	10	6–8	7.1	7
7	4	10–12	11	11	7–9	8.3	8.5
8	12	11–15	12.3	12	8–11	9	9
9	2	14	14	14	10	10	10
10	5	14–15	14.4	14	10–11	10.2	10
11	8	15–18	16.4	16	11–13	11.9	12
12	8	15–20	16.8	16.5	11–15	12.8	12.5
13	5	16–19	17.4	17	12–14	13	13
14	10	18–20	18.7	19	13–18	14.6	14
15	12	18–23	20.3	20.5	14–18	16.1	16
16	10	19–22	20.7	21	15–18	16.5	16
17	3	19–22	20.7	21	15–18	16.7	17
18	4	22–24	22.8	22.5	17–18	17.8	18

EN – English language, CRO – Croatian language, N – number of samples

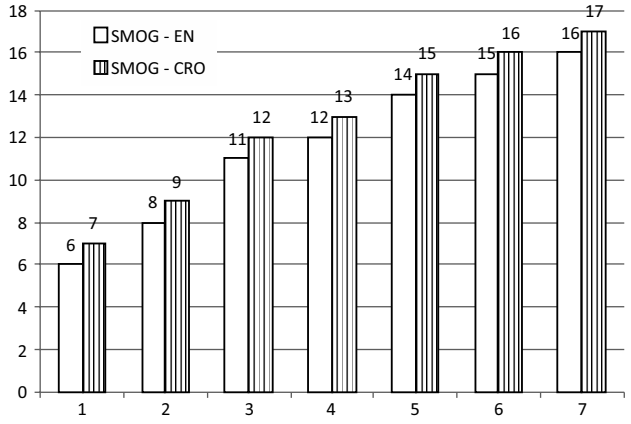

Fig. 1. Values of SMOG scores for English and Croatian language for selected seven scores found in more than five samples with 30 sentences. EN – English language, CRO – Croatian language.

number in French. On the other hand, the percentage of polysyllabic words found in this study, as could be calculated from the tables presented, was 16.8% for the total of 90 samples in English language but it varied among the four sets, with 8.6% for books of fiction, 19.9% for feature articles, 14.6% for popular science books, and 23.3% for research articles. Also, although the number of polysyllabic words for the total of 90 samples in Croatian language was 39.5% or 2.45 times greater than in English, the percentages also varied among the four sets, with 30.2%, 40.6%, 38%, and 49%, being greater than in English by 3.5, 2.0, 2.6, and 2.1 times in the four sets, respectively. The complexity of a language structure cannot, therefore, be determined from the simplest form of writing without biased conclusions. Contreras et al.[28] also provided both the formulas for converting the scores among the three languages and a table with estimated scores across the three languages. However, the table shows e.g. the score of 10 for English corresponding to rounded scores of: a) 18 in Spanish and 15 in French when converting from English; b) 17 in Spanish and 14 in French when converting from Spanish; and 19 in Spanish and 15 in French when converting from French. So, this 'ready-to-use conversion table' to be used for practical purposes, as the authors suggested, seems a bit confusing as an illustration of the equivalence of readability levels among the three languages. Çepni et al.[29] explained that the differences between English and Turkish language in the sentence length and number of polysyllabic words required adaptation of SMOG formula for Turkish language. They then defined the concept of »polysyllabic words« in Turkish language as those with four or more syllables, providing a readability formula for Turkish language that could be summarized in the form of: SMOG reading age $=5+\sqrt{4^+}$ syllables. They were not using the original McLaughlin's formula, where the coefficient to add is 3, but a revised one found in Wellington's book on science textbooks[31], where the coefficient of 8 is added to indicate the reading age of American readers, and not the years of education. The formula containing a coefficient 8, to indicate the reading age, and based on research at the Nottingham University, is found in the online calculator available since 2009 at the website of the The National Institute of Adult Continuing Education (NIACE), UK[32], as opposed to the online calculator with the original McLaughlin's formula[27]. Çepni et al.[29] mentioned that they based their adaptation of

the formula also on results they obtained by testing six science textbooks for primary school on pupils, using a Cloze procedure, but the number of primary school pupils tested is not consistent throughout the paper, and since the methodology in their paper is extremely short and lacking specific information, it is not exactly clear how the test itself was done. Also, they are pretty vague about the exact procedure for SMOG formula adaptation with »it is thought that 5 value should be supplemented to this square root value in order to reach more valid values«[29]. Kondilis et al.[30] described the readability levels of 70 health pamphlets in Greek language, and stated that Greek language is generally more complex with longer words and syllables than most other Romance languages. They themselves did not develop the SMOG formula used in their study, but referred to a software developed by the Centre for Greek Language in Thessaloniki, Greece, version 2000, based on 10 formula trials and 32 separate validation studies performed between 1999 and 2001. The reference is given for the Centre but the Centre website[34], apart from being in Greek language, does not seem to reveal any information either on the procedure of the formula adaptation for Greek language or on the actual equation used. Kondilis et al. even emphasize in their paper that the validity of readability formulas for Greek language »needs to be further assessed if these computerized tools developed by the Centre for Greek language studies are to be considered externally valid«[30].

Studies have shown that the scores of readability formulas correlate well with the comprehension difficulty as measured by reading tests[20]. Reading tests, such as WRAT and REALM, frequently used for speakers of English language, cannot be readily used for speakers of other languages even when translated[7], although some attempts have been made for different languages. The primary argument against translation of a reading test into Croatian language is orthographic transparency of the Croatian language, where words are pronounced without uncertainties or complex rules because of the sound-letter correspondences, and the proper pronunciation of a word would not automatically indicate comprehension of the word read.

A frequently used test of functional health literacy in adults, TOFHLA[10] and its shorter version S-TOFHLA[11], have been designed in the first place for English language, with passages of text to be included in the test used from actual hospital materials, and well defined as to their readability levels. The TOFHLA consists of a 50-item reading comprehension and 17-item numerical ability test, taking up to 22 minutes to administer. It shows good correlation with the WRAT revised version and the REALM reading tests. The three reading comprehension test passages have readability levels of 4.3, 10.4, and 19.5 according to the Gunning Fog readability formula, indicating school grade level, and the numeracy items have the average readability of 9.4. The shorter version, S-TOFHLA, taking only 12 minutes to complete, retained two passages for reading comprehension test, with 4th and 10th grade difficulty level, and four numeracy items, for which the authors do not provide readability scores. However, they describe the reliability of the numeracy items as only modest, with REALM correlation lower, and even consider removal of the numeracy items altogether suggesting that reading comprehension passages alone may prove useful as a screening instrument to identify patients with very limited reading ability. Thus, reading passages alone would be seen as a short clinical screening assessment of patients' literacy, even shorter than the presented S-TOFHLA, which they called »probably the shortest instrument possible that still provides a complete assessment of functional health literacy«[11].

To the best of our knowledge, and as informed by Connor et al.[19], currently the validated S-TOFHLAs only exist in a limited number of languages. It has been translated from English and validated into Spanish[11], Hebrew[12], Brazilian Portuguese[13], Serbian[15], Chinese (Cantonese)[16], Mandarin[17], and Korean[18]. Literature review shows that the first translation and/or cultural adaptation of TOFHLA, i.e. its short version S-TOFHLA, is a Hebrew version from 2007, which the authors called HHLT (The Hebrew Health Literacy Test)[12]. They described in detail the methodology used showing that they adapted the original questionnaire to a great extent to accommodate for the Israeli health system, language, and culture, and even the scoring of the HHLT was not based on the coding system of the S-TOFHLA since the two tools differed greatly. The authors did not mention readability levels of the texts used for the HHLT test. Tang et al. reported in 2008 to have used a Chinese version of S-TOFHLA, which they named C-S-TOFHLA, on 149 patients with diabetes mellitus[16]. However, no word-for-word translation was possible and the reading comprehension sections had to be rewritten[19], as well as in the Mandarin[17] and Korean[18] versions. Furthermore, Pan et al., who developed Health Literacy Scale for Chinese-speaking adults in Taiwan argued that not only are English and Chinese two totally different language systems with visible differences in terms of word composition, verbal expression (grammar system), and phonetics, but these »differences could also be exacerbated when terms or words were given and been accepted by others with additional meaning or metaphor that originated from a shared living experience«[35]. The authors of the Korean S-TOFHLA even concluded that their translated instrument »may not be fruitful in future research endeavors because of its poor performance with questionable validity«[18]. Jovic-Vranes et al.[14] used the TOFHLA instrument to assess functional health literacy among 120 primary healthcare patients in Belgrade, Serbia. They described briefly the translation and cultural adaptation process, mentioning the readability levels of the passages used for the test but only for the original they translated from, and not for the test they were actually using for their study population. They also reported that their Serbian version is the first TOFHLA translation outside the English-Spanish speaking countries, thus not acknowledging the Hebrew and Chinese versions developed on its short version some time earlier. Two years later, they reported having used the short version themselves, as it takes less time than the longer version, to determine health literacy levels of 1,500 primary healthcare patients[15]. Carthery-Goulart et al. – actually a group of 22 authors – used a

Brazilian version of S-TOFHLA on 312 healthy participants and reported that the English and Spanish versions of S-TOFHLA were translated and adapted to the Brazilian reality, especially the reading comprehension texts as to convey information about the Brazilian health system[13]. They did not report on the readability levels of the texts. Connor et al.[19] described the translation and cultural adaptation of S-TOFHLA into German, Italian, and French language. They provided a translation as close as possible to the original S-TOFHLA to make comparisons between countries possible but with some minor changes implemented due to differences in the Swiss healthcare system. They also reported that similarly to the original English version, all three translated versions of the S-TOFHLA consisted of three parts – two prose passages with a total of 36 cloze items and 1 numeracy section consisting of 4 numeracy items. However, neither these authors mentioned the readability levels of the passages they used for their tool of health literacy assessment. It is our opinion that, if the equivalence among the original health literacy assessment tool and the translated versions is to be kept, for the TOFHLA results to be comparable across countries and languages, the readability levels in the translated versions need to correspond to the levels found in the original. Any health literacy test using prose passages needs to have the text difficulty levels specified, as the original TOFHLA had with readability scores. Otherwise, we would not know which yardstick exactly we are using to measure such an important concept as health literacy, and all studies investigating the issue have reported that »literacy is a stronger predictor of an individual's health status than income, employment status, education level, and racial or ethnic group«[36]. For instance, years of schooling completed alone is an inaccurate indicator of educational attainment, because it merely signifies education attempted rather than attained[37].

It should be noted that readability formulas have often been criticized by the opponents of this approach to text difficulty assessment[20], disregarding warnings of their developers even early in the history of formulas development that these should be used only as rough guides, and not precise and exact values, to quickly estimate a text difficulty[38]. Those who criticize the formulas, often mentioning that only superficial language features are taken into account, such as sentence and/or word length, may have overlooked the facts deeply rooted in the linguistic patterns observed throughout time by studies into the language statistics. For example, word length is associated with word frequency, with the shortest words being the most frequent[39]. Chall, an author of a readability formula, even reported that vocabulary difficulty is the strongest predictor of text difficulty[40], and this finding is reflected in the SMOG formula, whose author stated that »a count of polysyllables is at least as valid as any other index of word difficulty, and it is certainly the easiest count to make«. He also explains that »the longer it takes to locate a word's meaning, the more likely it is that the preceding context will be lost beyond recall; thus word length, like sentence length, is an index of difficulty due to limitations of immediate memory«. Furthermore, he argues that those readability formulas that have been adequately validated actually predict comprehensibility. He considers his SMOG formula, although »laughably simple, in fact more valid than previous readability formulas«[24]. McLaughlin's observations that text difficulty in both the semantic and syntactic sense is associated with storage in short-term memory[24], have been confirmed by more recent studies that show how automaticity in word recognition is necessary for fluent reading, where the attention capacity is made free for comprehension. Fluent reading leads to reading precision, and the text is then accurately and precisely interpreted[39]. It is no surprise then that readability formulas show high correlation with comprehension tests performed on respondents. Studies into frequency of words have shown not only that the most frequent words are shorter, but also that people use more often some words, they recognize and learn them more quickly, understand them more easily, and even prefer them over others[41]. Fry, another author of a readability formula, found that the most frequent 100 words cover almost 50% of a text in English language[42]. This core vocabulary was the starting point for suggesting a readability formula for Croatian language in the previous studies[43,44]. The non-validated SMOG formula used in those studies only redefined the concept of polysyllabic words for Croatian language as words with four or more syllables, based on the difference observed in the syllable count of 100 most frequent words between English and Croatian language. The present study, based on analysis of almost 100,000 running words, provides a more precise modification of the original formula, which still should be seen as a rough estimate of text difficulty for Croatian language.

Having in mind that the first statistical analysis of language for the purpose of readability studies dates back to the late 19th century[45], and that the readability formulas have obviously stood the test of time, we hope that with the today's ease and speed of textual data collection and computer-assisted analysis of large textual corpora, we may expect further advances in the development of readability formulas for different languages. We also hope the methodology described in this paper helps in such developments, and that results of this study are put into practice soon for more effective healthcare communication and patient education, but also for development of a health literacy assessment tool in Croatian language.

REFERENCES

1. STEWART M, Can Med Assoc J, 152 (1995) 1423. — 2. KAPLAN RM, Am J Prev Med, 26 (2004) 81. DOI: 10.1016/j.amepre.2003.09.008. — 3. ROTER DL, HALL JA, Doctors talking with patients/patients talking with doctors (Auburn House, Westport/ London, 1992). — 4. WAITZKIN H, Changing patient-physician relationship in the changing health-policy environment. In: BIRD CE, CONRAD P, FREMONT AM (Eds) Handbook of medical sociology (Prentice Hall, New Jersey, 2000). — 5. MEAD N, BOWER P, Soc Sci Med, 51 (2000) 1087. DOI: 10.1016/S0277-

9536(00)00098-8. — 6. DOAK CC, DOAK LG, ROOT JH, Teaching patients with low literacy skills (JB Lippincott Company, Philadelphia, 1996). — 7. INSTITUTE OF MEDICINE, Health literacy: a prescription to end confusion (National Academies Press, Washington, 2004). — 8. AD HOC COMMITTEE ON HEALTH LITERACY, JAMA, 281 (1999) 552. — 9. OSBORNE H, Health Literacy from A to Z: Practical Ways to Communicate Your Health Message (Jones&Bartlett Publishers, Sudbury, 2005). — 10. PARKER RM, BAKER DW, WILLIAMS MV, NURSS JR, J Gen Intern Med, 10 (1995) 537. — 11. BAKER DW, WILLIAMS MV, PARKER RM, GAZMARARIAN JA, NURSS J, Patient Educ Couns, 38 (1999) 33. — 12. BARON-EPEL O, BALIN L, DANIELY Z, EIDELMAN S, Patient Educ Couns, 67 (2007) 235. DOI:10.1016/j.pec.2007.02.005. — 13. CARTHERY-GOULART MT, ANGHINAH R, AREZA-FEGYVERES R, SANTORO BAHIA V, DOZZI BRUCKI SM, DAMIN A, FORMIGONI AP, FROTA N, GUARIGLIA C, JACINTO AF, KATO EM, PIMA EP, MANSUR L, MOREIRA D, NOBREGA A, PORTO CS, SENAHA MLH, DA SILVA MNM, SMID J, SOUZA-TALARICO JN, RADANOVIC M, NITRINI R, Rev Saude Publica, 43 (2009) 631. DOI: 10.1590/S0034-89102009005000031. — 14. JOVIC-VRANES A, BJEGOVIC-MIKANOVIC V, MARINKOVIC J, J Public Health, 31 (2009) 490. DOI: 10.1093/pubmed/fdp049. — 15. JOVIC-VRANES A, BJEGOVIC-MIKANOVIC V, MARINKOVIC J, KOCEV N, Int J Public Health, 56 (2011) 201. DOI: 10.1007/s00038-010-0181-0. — 16. TANG YH, PANG SMC, CHAN MF, YEUNG GSP, YEUNG VTF, J Adv Nurs, 62 (2008) 74. DOI: 10.1111/j.1365-2648.2007.04526.x. — 17. TSAI TI, LEE SYD, TSAI YW, KUO KN, J Health Commun, 16 (2010) 50. DOI: 10.1080/10810730.2010.529488. — 18. HAN HR, KIM J, KIM M, KIM K, J Immigr Minor Health, 13 (2011) 253. DOI: 10.1007/s10903-010-9366-0. — 19. CONNOR M, MANTWILL S, SCHULZ PJ, Patient Educ Couns, 90 (2013) 12. DOI: 10.1016/j.pec.2012.08.018. — 20. DuBAY WH, The principles of readability (Impact Information, Costa Mesa, 2004). — 21. KLARE GR, The measurement of readability (Iowa State University Press, Ames, 1963). — 22. HARGIS G, HERNANDEZ AK, HUGHES P, RAMAKER J, ROUILLER S, WILDE E, Developing quality technical information: a handbook for writers and editors (Prentice Hall, Upper Saddle River, 1998). — 23. McLAUGHLIN GH, J Reading, 12 (1969) 639. — 24. McLAUGHLIN GH, Instr Sci 2 (1974) 367. DOI: 10.1007/BF00123459. — 25. GUNNING R, The technique of clear writing (McGraw-Hill, New York, 1952). — 26. SETimes, accessed 02.01.2011. Available from: URL: www.setimes.com. — 27. Words Count Readability, accessed 02.01.2011. Available from: URL: www.wordscount.info/wc/jsp/clear/analyze_readability.jsp. — 28. CONTRERAS A, GARCIA-ALONSO R, ECHENIQUE M, DAYE-CONTRERAS F, J Health Commun 4 (1999) 21. DOI: 10.1080/108107399127066. — 29. ÇEPNI S, GÖKDERE M, KÜÇÜK, M, Energy Education Science and Technology, 10 (2002) 49. — 30. KONDILIS BK, AKRIVOS PD, SARDI TA, SOTERIADES ES, FALAGAS ME, Public Health, 124 (2010) 547. DOI: 10.1016/j.puhe.2010.05.010. — 31. WELLINGTON J, Secondary science, contemporary issues and pratical approaches (PLC, London, 1994). — 32. NIACE SMOG Calculator, accessed 05.03.2013. Available from: URL: www.niace.org.uk/misc/SMOG-calculator/smogcalc.php#. — 33. TAYLOR WL, Journalism Q 30 (1953) 415. — 34. CENTRE FOR GREEK LANGUAGE STUDIES, Readability formulas for Greek text [in Greek], accessed 15.01.2013. Available from: URL: www.greek-language.gr/greekLang/modern_greek/foreign/tools/readability/index.html. — 35. PAN FC, SU CL, CHEN CH, International Journal of Biological and Life Sciences, 6 (2010) 171. — 36. WEISS BD, Health Literacy: A Manual for Clinicians (American Medical Association, Chicago, 2003). — 37. BAKER DW, PARKER RM, WILLIAMS MV, CLARK WS, NURSS J, Am J Public Health, 87 (1997) 1027. — 38. KLARE GR, ROWE PP, St JOHN MG, STOLUROW LM, Reading Research Quarterly, 4 (1969) 550. — 39. ALDERSON JC, Assessing Reading (Cambridge University Press, Cambridge, 2000). — 40. CHALL JS, Readability – an appraisal of research and application (Ohio State University, Columbus, 1958). — 41. KLARE GR, Elementary English, 45 (1968) 12. — 42. FRY EB, KRESS JE, FOUNTOUKIDIS DL, The reading teacher's book of lists (The Centre for Applied Research in Education, New York, 1993). — 43. KUSEC S, MASTILICA M, PAVLEKOVIC G, KOVACIC L, Stud Health Technol Inform, 90 (2002) 128. DOI: 10.3233/978-1-60750-934-9-128. — 44. KUŠEC S, OREŠKOVIĆ S, ŠKEGRO M, KOROLIJA D, BUŠIĆ Ž, HORŽIĆ M, Patient Educ Couns, 60 (2006) 294. DOI: 10.1016/j.pec.2005.10.009. — 45. SHERMAN AL, Analytics of literature: a manual for the objective study of English prose and poetry (Ginn&Co., Boston, 1893).

S. Brangan

University of Zagreb, School of Medicine, »Andrija Štampar« School of Public Health, Department of Educational Technology, Rockefellerova 4, 10000 Zagreb, Croatia
e-mail: skusec@snz.hr

RAZVOJ FORMULE ČITKOSTI SMOG-Cro ZA ZDRAVSTVENU KOMUNIKACIJU I EDUKACIJU PACIJENATA

S A Ž E T A K

Učinkovita komunikacija pozitivno utječe na zadovoljstvo pacijenata, njihovu suradljivost i ishode liječenja, ujedno smanjujući troškove zdravstvene zaštite. Pisane informacije za pacijente moraju biti sukladne razini zdravstvene pismenosti onih kojima su namijenjene. Formule čitkosti dobro koreliraju s testovima čitanja i razumijevanja, ali se smatraju jednostavnijom i bržom metodom procjene težine teksta. Formula čitkosti SMOG napravljena je za engleski jezik te se mora modificirati ako se koristi za tekstove na nekom drugom jeziku. Cilj ovog istraživanja bio je razviti formulu čitkosti temeljenu na formuli SMOG, a koja će se moći koristiti za procjenu težine teksta pisanih materijala za pacijente na hrvatskom jeziku. Kontrastivna analiza engleskog i hrvatskog jezika na korpusu od skoro 100.000 riječi pokazala je jasne jezične razlike u broju višesložnih riječi. Nova formula, nazvana SMOG-Cro, prikazana je kao jednadžba: SMOG-Cro $=2+\sqrt{4^{+}}$ sloga, čiji rezultat pokazuje koliko je godina školovanja potrebno nekoj osobi da bi mogla razumjeti određeni tekst. Metodologija prikazana u ovom radu mogla bi pomoći pri razvoju formula čitkosti za druge jezike. Nadamo se da će rezultati ovog istraživanja uskoro biti i praktično primijenjeni za što uspješniju zdravstvenu komunikaciju i edukaciju pacijenata, ali i za razvoj instrumenta za ocjenu zdravstvene pismenosti na hrvatskom jeziku.