

# Simplified Readability Metric Drives Improvement of Radiology Reports: an Experiment on Ultrasound Reports at a Pediatric Hospital

Wei Chen<sup>1</sup> · Claire Durkin<sup>2</sup> · Yungui Huang<sup>1</sup> · Brent Adler<sup>3</sup> · Steve Rust<sup>1</sup> · Simon Lin<sup>1</sup>

Published online: 8 May 2017

© Society for Imaging Informatics in Medicine 2017

**Abstract** Highly complex medical documents, including ultrasound reports, are greatly mismatched with patient literacy levels. While improving radiology reports for readability is a longstanding concern, few articles objectively measure the effectiveness of physician training for readability improvement. We hypothesized that writing styles may be evaluated using an objective two-dimensional measure and writing training could improve the writing styles of radiologists. To test it, a simplified “grade vs. length” readability metric is developed based on results from factor analysis of ten readability metrics applied to more than 500,000 radiology reports. To test the short-term effectiveness of a writing

workshop, we measured the writing style improvement before and after the training. Statistically significant writing style improvement occurred as a result of the training. Although the degree of improvement varied for different measures, it is evident that targeted training could provide potential benefits to improve readability due to our statistically significant results. The simplified grade vs. length metric enables future clinical decision support systems to quantitatively guide physicians to improve writing styles through writing workshops.

**Keywords** Readability metrics · Radiology reports · Ultrasound reports · Writing styles · Factor analysis

✉ Wei Chen  
wei.chen@nationwidechildrens.org

Claire Durkin  
claire.durkin@osumc.edu

Yungui Huang  
yungui.huang@nationwidechildrens.org

Brent Adler  
brendan.boyle@nationwidechildrens.org

Steve Rust  
steve.rust@nationwidechildrens.org

Simon Lin  
simon.lin@nationwidechildrens.org

<sup>1</sup> Research Information Solutions and Innovation, Nationwide Children’s Hospital, 575 Children’s Crossroad, Columbus, OH 43215, USA

<sup>2</sup> College of Medicine, The Ohio State University, Columbus, OH, USA

<sup>3</sup> Department of Radiology, Nationwide Children’s Hospital, Columbus, OH, USA

## Introduction

The radiology report, one important type of complex medical document, is the primary record of the interpretation of imaging procedures. Classically, this document serves as the basis of communication between radiologists and ordering physicians. However, a recent regulatory push is to provide patients access to health records (including radiology reports) with the aim of keeping patients and patients’ family better informed. Ultimately, this information is intended to empower patients to make informed decisions and to improve healthcare outcomes [1]. To achieve the aims of open access, making medical reports comprehensible to patients is critical [2, 3].

Improved accessibility does not necessarily lead to improved understanding. For most patients with little or no medical literacy, understanding a typical medical report can be challenging [3, 4]. This mismatch between relatively low patient literacy levels and highly complex medical documents, including radiology reports, is a longstanding, recognized problem [5, 6]. While previous studies found many medical documents and educational materials were written at college

level or above [7, 8], the US census indicates the average US adult reads at the 7th to 9th grade level [4, 6].

Readability measures have been developed in evaluating the literacy level of medical documents. Reading grade-level metrics (e.g., Flesch-Kincaid, Gunning Fog, Coleman-Liau, and automated readability index) are effective for characterizing the reading grade levels of medical education materials [9], clinical trial results [10], patient consent [11], and cancer screening announcement [12] as well as radiology reports [13]. Besides, more sophisticated natural language processing (NLP) metrics have also been implemented to evaluate readability. For example, part-of-speech (POS) tagging and function-word extraction are useful in assessing a medical report grammatically [13, 14]. However, POS tagging is likely to be error prone when parsing text full of abbreviations and numbers, which are typical of radiology reports [15]. Such errors can propagate and falsely indicate sentence levels up to 2.5 times [16]. Among all POS errors, noun phrase errors are particularly predominant (7% error rate), contributing greatly to misunderstanding medical terms in radiology reports, which are primarily noun phrases [15]. Therefore, traditional lexicon-based readability measurements are still overall more reliable than sophisticated NLP metrics for readability assessment.

Despite the feasibility demonstrated in previous research [17], two major limitations constrain improving radiologists' writing for better patient comprehension. First, most study analyses were limited to a relative small number of a few hundred documents. Only one study used a large dataset (about one million records). However, the documents in this study (doctors' referral letters and MedlinePlus articles [18]) were less complex than ultrasound reports. Second, although a 2014 study discovered some demand from patients to provide radiologists with training to improve the written radiology reports [19], little effort has materialized. Empirical evidence is needed to verify that "radiologists' reports can substantially improve with training and rapid feedback" [19].

This study is the first to design a user experiment to assess the effects of training aimed at improving radiological writing styles. Unlike previous studies that relied on direct readability metrics or NLP measurements, we introduced an interpretable metric, "grade vs. length" writing style, to quantitatively assess writing styles before and after a planned training. The grade vs. length metric was derived from factor analysis of 10 previously reported readability indices. Finally, we trained 14 radiologists through two rounds of user experiments for written comparisons. Changes in writing style of each participating radiologist were quantified, visualized, and statistically tested to provide firsthand evidence on training effectiveness.

## Materials and Methods

### Data

In this study, ultrasound reports were used as a subset of a larger corpus of radiology reports. We only included ultrasound reports from children evaluated between January 2012 and December 2014. Radiology reports from our hospital's electronic medical record (EMR) system through Epic (Epic Systems, Inc.) were extracted resulting in about a half million eligible ultrasound reports.

Preprocessing of ultrasound reports was necessary to cull some sections, which are not relevant for patient understanding. The ultrasound report structure of most US hospital systems include the following sections: *patient information*, *technique*, *contrast*, *comparison*, *findings*, and *impressions* [20]. Our study focused on the ultrasound report "impressions" section, which summarizes the ultrasound results of greatest interest to patients (e.g., information about findings, diagnoses, and recommendations). We excluded other sections from our analysis since those sections are basically in support of the "impressions" section claims.

### Data Preparation

To extract the "impressions" section, a customized regular expression (regex)-parser was implemented. The regex-parser detects section headings as boundaries (e.g., FINDINGS and IMPRESSIONS; usually in plural form, in capitals, and followed by a colon), based on which the content in between can be extracted. A regex-parser is sufficient in our case because section boundary markers were systematically designed by our physician lab and were formulated as templates for constructing new reports. The section marker "IMPRESSIONS:" was unique throughout the text of an ultrasound report.

Text inputs that are irrelevant for readability analysis in the "impressions" section were removed. Such inputs include placeholder characters and signature lines. The placeholder characters, which often appeared at the beginning of a sentence, are composed of markers such as "I." and "A.". Signature lines such as "I, <name of physician>, have supervised ..." were also removed also using a customized regex-parser.

### Readability Measures

Prior to characterizing the writing styles of radiologists at our hospital, we established 10 recognized readability measures (Table 1). Complex words (relating to measures 3 and 5) are defined as words with three or more syllables. Other measures are considered self-explanatory by their names and full definition.

**Table 1** Readability measures of ultrasound reports

ID	Measure	Type	Definition/formula
1	Number of Characters (NC)	Lexical	Length in character.
2	Number of Syllables (NS)	Lexical	Total number of syllables in a sentence.
3	Number of Complex Words (NCW)	Lexical	Number of words with 3 and more syllables.
4	Section Length (SL)	Lexical	Number of words in the “impression” section.
5	Percentage of Complex Words (PCW)	Lexical	NCW/SL.
6	Average Syllables per Word (ASW)	Lexical	NS/SL.
7	Flesch-Kincaid (FK)	Grade level	$0.39 * SL + 11.8 * NS / SL - 15.59$ .
8	Gunning Fog (GF)	Grade level	$0.4 * (SL + 100 * NCW / SL)$
9	Coleman-Liau (CL)	Grade level	$(5.89 * NC / SL) - (29.6 / SL) - 15.8$
10	Automated readability index (ARI)	Grade level	$4.71 * NC / SL + 0.5 * SL - 21.43$

Since not all reading grade-level formulas use the same type of lexical information for calculation, different readability measures in Table 1 may represent slightly different aspects of readability. For example, *Coleman Liau* and *automated readability index (ARI)* measures rely on the number of characters instead of number of syllables for calculation. Although measures 7–10 were calculated based on measures 1–6, measures 1–6 uniquely differ from the derived readability grade-level measures.

Previous studies utilized more sophisticated NLP measures including POS tags and parsing score of a sentence for readability evaluation [13–15, 21]. However, given that our ultrasound corpus embodies many incomplete sentences, missing and misplaced apostrophes, abbreviations, and spelling errors, our initial experiment found the use of these sophisticated NLP measures ineffective and subject to large margin of errors for processing. Such unstandardized text may be acceptable and appropriate for human-to-human communication, but poses serious challenges to accurate NLP processing.

### Factor Analysis and Writing Style

Factor analysis, a common dimensionality reduction technique, can reduce a set of variables to a smaller set of orthogonal components—called factors [22]—and is effective in reducing the dimensions of numeric data [23]. To determine if the 10 readability measures in Table 1 could be reduced to two variables as in a two-dimensional space, we applied factor analysis.

Factor analysis was adopted for two reasons. First, grouping by factor analysis allows us to select the most relevant readability measures to derive a writing style metric, a two-dimensional representation based on factor analysis results. A metric could be easily visualized and become more understandable by clinicians in a two-dimensional space than in a multi-dimensional space. Second, through factor analysis, we could pair down the number of measures so that a few selected metrics could be targeted in the radiology workshops.

We selected two factors that account for 91% of the variability in our data [24] and, based on correlations between these factors and the original 10 readability measures, divided the 10 readability measures into two primary groups: *Section Length* and *Composite Grade Level*. See the “Results” section for details.

### Physician Training for Improving Writing Styles

To evaluate the effectiveness of training for improving writing styles toward better comprehension, we selected two ultrasound cases and conducted two rounds of experiments with our 14 participating radiologists. The two randomly selected ultrasound cases included one relatively simple image and one relatively complex image, as determined by the head of the radiology department.

To compare the results before and after the training, we designed two rounds of experiment. In the first experiment, the 14 radiologists were asked to write the impression section for each of the two ultrasound cases. To ensure the results were representative of their normal writing style, we presented the task to the participants without the knowledge of the following training and second experiment.

Following the first experiment, the authors presented to the participating radiologists a 1-h workshop, which emphasized writing style characterization and effective radiology report writing such as simple sentence structure and brevity [19] and how to write more effectively without losing critical information [20]. The workshop encouraged the use of simpler words and phrases and less complex sentence structures to enhance readability. Also, during the 1-h workshop, radiologists received their own readability scores on the two ultrasound-report cases and a summary of how one radiologist’s writing style compared to another.

The second experiment followed within 2 weeks after the workshop. All participants were again asked to write the impression section on the same two cases using the principles learned during the training. Additionally, we asked

participants to fill out a survey after the second experiment. The results of the survey were purposed to facilitate the design of workshops in the future.

We tested each of the two cases for statistically significant changes in writing styles. Paired *T* tests, as a parametric measure, were used to gauge the significance of writing style change. Since we only have a relatively small group of 14 radiologists participating in the study, we also employed Wilcoxon signed-rank test, as a non-parametric alternative, to further confirm the results in the *T* test.

## Results

### Writing Styles Quantified by Section Length and Composite Grade Level

To develop the writing style metric, we used all “impressions” sections to date written by 25 radiologists at our hospital between 2012 and 2014, which included about 500,000 narratives. Ten readability measures were calculated for each narrative, and a statistical summary of the results is reported in Table 2 along with factor analysis results. In Table 2, the median value of each measure, its standard deviation, and the interquartile range are reported.

Factor analysis resulted in two factors which were later used to construct the two dimensional writing style metric. Factor 1, primarily representing and correlating with the length of the section, explained 55% of total variance in the data, while factor 2, primarily representing readability measures, explained an additional 36% of variance. In total, two factors explained 91% of the variance indicating these extracted factors were good surrogates of the original 10 variables.

Readability loadings for factors 1 and 2 are reported in the last two columns of Table 2. As noted in the bolded measure loadings, factor 1 is driven primarily by measures 1–4, and factor 2 is driven primary by measures 5–10.

In order to simplify the interpretation of our statistical analysis results, we decided to base the analysis on two characteristic measures, one representing measures 1–4 and the other representing measures 5–10. Measure 4, Section Length (SL), was selected to represent the first group and Composite Grade Level (CGL) was derived to represent the second group where CGL is calculated as follows:

$$\text{CGL} = (\text{FK} + \text{GF} + \text{CL} + \text{ARI})/4$$

where Flesch-Kincaid (FK), Gunning Fog (GF), Coleman-Liau (CL), and automated readability index (ARI) are specific measures 7–10, respectively.

The correlation between variables in each group and the corresponding characteristic measure showed that the two characteristic measures are good representations. The correlations between SL and Number of Characters (NC), Number of Syllables (NS), and Number of Complex Words (NCW) were 0.99, 0.92, and 0.98, respectively (all  $p < 0.01$ ). The correlations between CGL and Percentage of Complex Words (PCW), Average Syllables per Word (ASW), FK, GF, CL, and ARI were 0.81, 0.84, 0.95, 0.90, 0.92, and 0.95, respectively (all  $p < 0.01$ ). However, the correlation between two characteristic measures SL and CGL is only  $-0.05$  ( $p = 0.01$ ). Although the  $p$  value of 0.01 implies a statistically significant correlation, it is probably due to the large sample size we have and the small correlation coefficient indicates that the correlation is of only small practical significance.

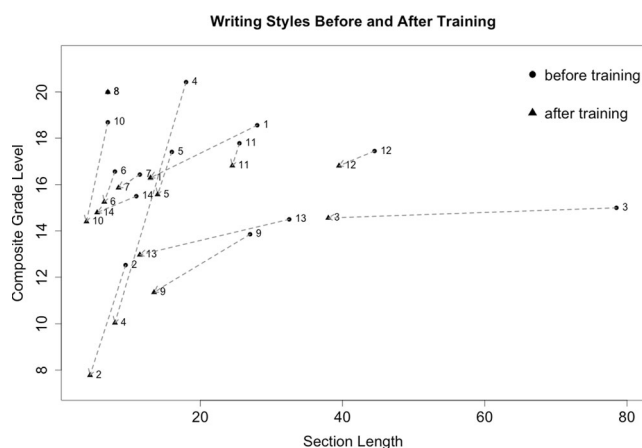
Finally, a coordinate pair ( $x, y$ ) where  $x$  is SL and  $y$  is CGL is representative of a physician's writing style and is calculated as the average of the two writing style analyses on each of the reports written by that physician. The difference in writing styles among physicians can therefore be characterized by the distances between the dots on a grade vs. length plot. Further, the magnitude of change before and after training can also be indicated by the direction and length of the movement of the same dot representing the same physician (Fig. 1).

**Table 2** Readability statistics and factor analysis results on 500,000 ultrasound radiological reports

ID	Measure name	Median (IQR <sup>a</sup> )	Standard deviation	Factor 1 <sup>b</sup>	Factor 2 <sup>b</sup>
1	<i>Number of Characters (NC)</i>	147(85, 244)	144	<i>0.996</i>	0.035
2	<i>Number of Syllables (NS)</i>	45(26, 75)	44	<i>0.988</i>	0.107
3	<i>Number of Complex Words (NCW)</i>	7(4, 11)	7	<i>0.860</i>	0.371
4	<i>Section Length (SL)</i>	21 (12, 36)	22	<i>0.976</i>	−0.129
5	<i>Percentage of Complex Words (PCW)</i>	0.32(0.25, 0.39)	0.13	−0.139	<i>0.915</i>
6	<i>Average Syllables per Word (ASW)</i>	2.13(1.95, 2.35)	0.39	−0.215	<i>0.939</i>
7	<i>Flesch-Kincaid (FK)</i>	14(12, 17)	5	0.185	<i>0.937</i>
8	<i>Gunning Fog (GF)</i>	17(14, 20)	5	0.213	<i>0.895</i>
9	<i>Coleman Liau (CL)</i>	21(18, 24)	5	0.131	<i>0.885</i>
10	<i>Automated readability index (ARI)</i>	16(14,19)	4	0.309	<i>0.876</i>

<sup>a</sup> IQR, interquartile range is defined as between the 25th and 75th percentile and the median is the 50th percentile.

<sup>b</sup> Italicized numbers indicate the larger value of the loadings for the two factors of each variable. The variable is best represented by the factor with the larger loading value.



**Fig. 1** Writing style change before and after training for radiologists

### Writing Style Improvements

Each radiologist's writing style is represented as a dot in Fig. 1 with the  $x$ -axis representing the section length and the  $y$ -axis representing the composite grade level. Improvements in writing style are suggested by the direction and the distance of change between the before-training symbols (dots) and the after-training symbols (triangles) for each radiologist. The numbers by the symbols are the unique identifiers for each radiologist across experiments to maintain anonymity.

As expected, most dots moved left and down (i.e., shorter section length and lower composite grade level after training, respectively). The length of the arrow indicates the degree of change (greatest for radiologists no. 3 in section length and no. 4 in grade level; Fig. 1). While radiologist no. 8 did not make any changes between experiments, this radiologist's style is already characterized by concise writing limiting room for change in the second experiment although grade level remained high. Generally, those who wrote briefer sections and at lower grade level in the first round would understandably have less room for improvement.

Both parametric Student  $T$  tests and Wilcoxon signed-rank tests showed statistical significance in writing style change on both dimensions of SL and CGL (Table 3). After Bonferroni correction, the  $p$  values for both tests are still significant. The Bonferroni correction adjusts the significance level by

**Table 3** Statistical testing of writing style change for results before and after training

Paired variable (before minus after)	Paired differences mean (95% confidence interval)	$T$ test $p$ value (2- tailed)	Wilcoxon signed-rank $p$ value (2-tailed)
Section length	8.75 (1.36, 16.14)	0.024	0.006
Composite grade level	2.01 (0.34, 3.69)	0.022	0.007

dividing 0.05 by the number of individual tests. In this case, the significance level for each individual hypothesis is  $0.05/2 = 0.025$ . The initial position of radiologist no. 3 represented a potential outlier for our analysis; but the Wilcoxon signed-rank test, which is robust to such potential outliers, confirmed the results from the  $T$  test.

Before the training, the average section length was 23 words and the average reading grade level was grade 16. To read and understand text at this level, a graduate level of education or above is needed. After the training, the average section length decreased to 15 words and the average reading grade level dropped to college level (grade 14). A 1-h workshop emphasizing writing style and readability saw a 35% improvement in brevity (section length reduction) and 13% improvement in lowering grade level (grade level reduction), both of which are statistically significant.

### Survey of Radiologists' Attitudes Toward Writing Training

When surveyed after the training workshop, all participating radiologists agreed that one's writing styles could be changed (Table 4). A majority (57%) believed that a change of writing style need not compromise accuracy. Also, most (71%) believed writing styles for communication with both physicians and patients could be optimized. Finally, most (64%) believed that training in improved writing styles could help improve patient comprehension.

### Discussion

To evaluate the writing styles by radiologists, we developed a two-dimensional metric which was easy to visualize and interpret. Two writing samples from 14 radiologists were used to detect any writing style change due to a 1-h workshop on improving written readability.

Before the training, radiologists wrote reports at an advanced college level (grade 16) on average and dropped to an early college level (grade 14) after training. This statistically significant but small change observed in reducing the grade level suggests either significant resistance and/or further challenge to continue lowering the grade level of radiology reading materials while maintaining the critical information in the report. By comparison, section length is a more tangible measure and is relatively easier to improve (a 35% reduction in section length was achieved in this study). Such an observation, previously not available, suggested that ultrasound reports at our hospital were written at a level that was beyond the expected reading level of average patients (grade 7–9) and effective training may potentially narrow this gap based on the available amount of the contrast experiment data that were collected in this research [19].



**Table 4** Physician attitudes regarding writing style and training for radiologists

Question	Yes	No
Do you think one's writing style can be changed?	14 (100%)	0 (100%)
Do you think the change of writing style will eventually compromise accuracy?	6 (43%)	8 (57%)
Do you think there is an optimal writing style for communication with physicians?	10 (71%)	4 (29%)
Do you think there is an optimal writing style for communication with patients?	10 (71%)	4 (29%)
Do you think some training in writing reports will help you improve your writing style to be more easily understandable by patients?	9 (64%)	5 (36%)

Unlike previous studies [25], we added section length as a separate metric for evaluating readability in combination with grade level. Although prior research on health literacy and readability indicated that shorter is “always” better for comprehension [19, 26], in radiology report writing, shortening is not always easy as one often needs to interpret and clarify findings and additional explanation tends to lengthen the passages. Therefore, it is still a meaningful yet non-trivial effort to improve brevity.

By comparison, our results also suggest that lowering the reading grade level may be potentially more challenging. To lower the reading grade level, we may need additional hours of professional training; or we may need to overcome scholarly reticence (43%; Table 4) to “dumbing down” the language for fear of compromising accuracy in these technical, medical reports; or both. Regardless, shortening section length is a viable goal for improving readability of ultrasound report in the patient population.

Our study was the first to carry out a contrast experiment to gauge the change in readability of ultrasound reports due to training radiologists in simpler writing styles. Despite the relatively limited number of samples collected from our participants, the writing style changes observed were statistically significant. In addition, responses from the radiologists in the post-experiment survey were a consensus that writing style could be improved and most agreed that writing workshops could be a useful tool in helping radiologists attain readability. Overall, our results supported our claim that writing styles of radiologists could potentially be improved through targeted training.

### Limitations

Our findings were subject to several limitations. First, traditional readability measures are designed to be focusing on lexical complexity rather than the comprehension of the text [27, 28]. Health literacy measures, such as the Test of

Functional Health Literacy in Adults [29], are alternatives that focus on comprehension (patient-centered). To implement such health literacy measures, however, would have required us to recruit patients for the study. Interacting with both ordering physicians and patients and designing two separate experiments was beyond the scope of our study.

Moreover, we have not explored the difference in the frequency of use between everyday language and medical language. Such frequency measures have not been previously explored to study medical literacy, but previous studies suggest that frequently used words are more comprehensible than less frequent ones such as medical jargons [30, 31]. However, research on which medical terms are used more frequently is scant. Therefore, incorporating frequency of use into studies is challenging [13].

Finally, observed training effect may be short-term due to our limited time and scope of training. We have not been able to collect empirical data big enough to determine the long-term effect of training. Since this study was performed at a single site, our methods also may need to be adjusted to generalize for other institutions. Also, given the innate differences in individuals' writing styles and learning curves, the degree of writing style change as a result of workshop training could vary (Fig. 1).

### Conclusions

Our study demonstrated the feasibility and effectiveness of conducting writing workshops to improve the readability of ultrasound reports at a level meant to communicate with potential non-clinical readers. Based on factor analysis of approximately 500,000 radiology reports, we characterized each radiologist's writing style using a two-dimensional representation based on composite reading grade level and section length. By conducting evaluations in writing styles before and after training, we could effectively measure the efficacy

of the training program. Despite the high reading complexity of ultrasound reports, the empirical evidence suggested a viable though challenging writing style improvement pathway for radiologists. Additional time in training may be warranted to affect even greater improvement. However, our 1-h training session was effective, efficient, and statistically significant in making improvements possible without affecting the primary message communicated in the report according to the head of the radiology department.

The clinical impact of using the simplified grade vs. length metric enables implementation in future clinical decision support systems, possibly as ambient quality measures for real-time feedback or in a dashboard for periodic physician review. Future research is needed to determine how the linguistic variables identified in this study influence patients' understanding of their conditions, utilization of the healthcare system, and ultimately, clinical outcomes. We encourage researchers to apply the methods used in this study to a broader range of clinical writings and to experiment with additional measures that account for reader comprehension.

**Acknowledgements** We would like to thank Richard Hoyt and Megan Reynolds from Nationwide Children's Hospital's Research Information Solutions and Innovation (RISI) center for their assistance with retrieving data from Epic. We also thank Tran Bourgeois for using REDCap (Research Electronic Data Capture) to set up user experiments, Katherine Strohman for project management, Dr. Jeffrey Hoffman for discussion of EPIC integration, and Dr. Michael Bruno from the Milton S. Hershey Medical Center at the Penn State College of Medicine for inspiring the study.

#### Compliance with Ethical Standards

**Funding** This study was supported by grant UL1TR001070 from the Center for Clinical and Translational Science at the Ohio State University to Dr. Brent Adler and MDSR Roessler Scholarship from the Ohio State University medical center to Claire Durkin.

**Competing Interests** The author's declare that they have no competing interests.

#### Reference

- Weiner SJ, Schwartz A, Sharma G, et al. Patient-centered decision making and health care outcomes: an observational study. *Annals of internal medicine* 2013;**158**(8):573–79
- Woods SS, Schwartz E, Tuepker A, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. *Journal of medical Internet research* 2013;**15**(3):e65
- Landro L Radiologists push for medical reports patients can understand. *Wall Street Journal* 2014
- Safeer RS, Keenan J. Health literacy: the gap between physicians and patients. *Am Fam Physician* 2005;**72**(3):463–68
- Abujudeh H, Pyatt RS, Bruno MA, et al. Radpeer peer review: relevance, use, concerns, challenges, and direction forward. *Journal of the American College of Radiology* 2014;**11**(9):899–904
- Kirsch IS Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey: ERIC, 1993
- Svider PF, Agarwal N, Choudhry OJ, et al. Readability assessment of online patient education materials from academic otolaryngology–head and neck surgery departments. *American journal of otolaryngology* 2013;**34**(1):31–35
- Eloy JA, Li S, Kasabwala K, et al. Readability assessment of patient education materials on major otolaryngology association websites. *Otolaryngology–Head and Neck Surgery* 2012;**147**(5):848–54
- Friedman DB, Hoffman-Goetz L. A systematic review of readability and comprehension instruments used for print and web-based cancer information. *Health Education & Behavior* 2006;**33**(3):352–73
- Friedman DB, Kim S-H, Tanner A, Bergeron CD, Foster C, General K. How are we communicating about clinical trials?: an assessment of the content and readability of recruitment resources. *Contemporary clinical trials* 2014;**38**(2):275–83
- Valentini M, Daniela D, Alonzo MCP, Lucisano G, Nicolucci A: Application of a readability score in informed consent forms for clinical studies. *Journal of Clinical Research & Bioethics* 2013: **2013**
- Okuhara T, Ishikawa H, Okada H, Kiuchi T. Readability, suitability and health content assessment of cancer screening announcements in municipal newspapers in Japan. *Asian Pacific journal of cancer prevention: APJCP* 2014;**16**(15):6719–27
- Zeng-Treitler Q, Goryachev S, Kim H, Keselman A, Rosendale D. Making texts in electronic health records comprehensible to consumers: a prototype translator. *AMIA Annu Symp Proc* 2007:846–50
- Leroy G, Helmreich S, Cowie JR, Miller T, Zheng W. Evaluating online health information: beyond readability formulas. *AMIA Annu Symp Proc* 2008:394–8
- Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assn* 2010;**17**(5):507–13
- Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. *Yearb Med Inform* 2008:128–44
- Osborne H.: *Health Literacy from A to Z*: Jones & Bartlett Publishers, 2012
- Applying multiple methods to assess the readability of a large corpus of medical documents. 14th World Congress on Medical and Health Informatics, MEDINFO 2013; 2013
- Bruno MA, Petscavage-Thomas JM, Mohr MJ, Bell SK, Brown SD. The “open letter”: radiologists' reports in the era of patient web portals. *Journal of the American College of Radiology* 2014;**11**(9):863–67
- Flanders AE, Lakhani P. Radiology reporting and communications: a look forward. *Neuroimaging Clinics of North America* 2012;**22**(3):477–96
- Penniston A, Harley E: Attempts to verify written english. *Proceedings of The Fourth International Conference on Computer Science and Software Engineering* 121–28, 2011
- Comrey AL, Lee HB. *A First Course in Factor Analysis*: Psychology Press, 2013
- Kline P *An easy guide to factor analysis*: Routledge, 2014.
- Cliff N. The eigenvalues-greater-than-one rule and the reliability of components. *Psychol Bull* 1988;**103**(2):276–79
- Hansberry DR, John A, John E, Agarwal N, Gonzales SF, Baker SR. A critical review of the readability of online patient education

- resources from RadiologyInfo. Org. American Journal of Roentgenology 2014;**202**(3):566–75
26. AHRQ: Health Literacy. Secondary Health Literacy 2013. <http://nnlm.gov/outreach/consumer/hlthlit.html>.
27. Leong EK, Ewing MT, Pitt LF. E-comprehension: evaluating B2B websites using readability formulae. Industrial Marketing Management 2002;**31**(2):125–31
28. Allen ED, Bernhardt EB, Berry MT, Demel M. Comprehension and text genre: an analysis of secondary school foreign language readers. The Modern Language Journal 1988;**72**(2):163–72
29. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults. Journal of general internal medicine 1995;**10**(10):537–41
30. Klare GR. The role of word frequency in readability. Elementary English 1968;**45**(1):12–22
31. Marks CB, Doctorow MJ, Wittrock MC. Word frequency and reading comprehension<sup>1</sup>. The Journal of Educational Research 1974;**67**(6):259–62