Gist Inference Scores predict gist memory for authentic patient education cancer texts

Mitchell Dandignac*, Christopher R. Wolfe

Department of Psychology, Miami University, Oxford, OH 45056, United States

ABSTRACT

Objectives: Develop a tool to evaluate and improve written medical communication to patients. Determine how effectively Gist Inference Scores (GIS) predict comprehension of patient education texts independently of health literacy. Explicate the text characteristics and psychological mechanism underlying GIS.

Methods: For study 1, a nationally representative sample of older women (N = 61) completed a fill-in-the-blank comprehension task on authentic National Cancer Institute (NCI) texts of varying GIS levels. In study 2, participants (N = 198) read NCI texts (high or low GIS) then recalled what they read.

Results: Study 1 showed that a higher percentage of different words yielding semantically similar sentence meaning were used to correctly fill-the-blanks on high GIS texts and there was no significant interaction with health literacy. In study 2, a greater proportion of decision-relevant information was recalled for high GIS texts.

Conclusions: GIS predicts the likelihood that readers will form gist representations of medical texts on free recall and fill-in-the-blank tasks. High GIS texts allow for more semantic flexibility to mentally represent the same meaning, and more strongly emphasizes gist rather than verbatim representations.

Practical implications: GIS provides medical communicators with an automated and user-friendly method to evaluate medical texts for their ability to convey the bottom-line meaning.

© 2020 Elsevier B.V. All rights reserved.

ARTICLE INFO

Article history:
Received 7 August 2019
Received in revised form 11 February 2020
Accepted 17 February 2020

Keywords:
Gist Inference Scores
Written communication
Comprehension
Fuzzy-Trace theory
Coh-Metrix

1. Introduction

Improving health literacy is crucial because low health literacy can impair a patient’s ability to absorb decision-critical information from medical professionals. However, relaying written medical information to patients, such as through webpages and brochures, is challenging because the knowledge gap between medical experts and non-experts can hinder communication. An important avenue for research is developing better tools to help medical professionals evaluate and improve their patient education texts (i.e. brochures and webpages) to improve communication. Digital discourse technologies offer medical professionals tools to automatically evaluate their patient education texts to predict how everyday people will comprehend them [1]. Traditional qualitative methods of text analysis require considerable resources and traditional computational methods, such as Flesch–Kincaid Grade Level (FKGL), provide a superficial analysis of readability that do not take advantage of emerging discourse technologies. Coh-Metrix is a freely available discourse technology (cohemetrix.com/tool) that provides data-rich computational analyses of texts (web sites, pamphlets, articles, etc.) at multiple levels including individual words, sentences, paragraphs, and the entire text [2]. Coh-Metrix computes over 100 linguistic variables across many analytical categories including: descriptive statistics, text “easability” principle components, referential cohesion, Latent Semantic Analysis, lexical diversity, connectives, situation model, syntactic complexity, syntactic pattern density, and word information [2].

Recent research has utilized Coh-Metrix variables to make predictions stemming from Fuzzy-Trace Theory (FTT), a theory of medical decision making [3–6]. FTT posits that people form multiple mental representations along a continuum from the superficial verbatim details to the fuzzier gist representation that conveys the bottom-line meaning [4]. Moreover, people prefer to make decisions using gist representations of the bottom-line meaning. This preference grows with expertise [7] and through cognitive development [8]. Importantly, patients may struggle to create gist representations of medical text (i.e. the merits of different treatment options) if the text emphasizes peripheral details because they lack the expertise to find meaning in those

* Corresponding author.
E-mail address: dandigm@miamioh.edu (M. Dandignac).

https://doi.org/10.1016/j.psc.2020.02.027
0738-3991 © 2020 Elsevier B.V. All rights reserved.
details. Medical practitioners may struggle to evaluate the comprehensibility of medical texts from a patient perspective because they have the expertise to extract meaning from those peripheral details. According to FTT, patients will make better medical decisions if they can form the preferred gist representations from medical text.

Guided by FTT, Gist Inference Scores (GIS) are a proximal index of the likelihood that readers will infer the bottom-line meaning of expository texts [9–11]. GIS is a formula integrating seven Coh-Metrix psycholinguistic variables (Fig. 1) that measures the extent to which a reader will generate meaningful inferences from a text. The Coh-Metrix variables assessing cohesion and abstract verb overlap are weighted positively, while verbatin verb overlap and word concreteness are weighted negatively (for readers who wish to calculate GIS scores for any text themselves, an Excel spreadsheet, ConvertCohMetrixToGIS_Scores.xlsx, that converts Coh-Metrix output to Gist Inference Scores has been included on the journal web site under supplemental materials; see [9] for technical details). Cohesion and abstract language facilitate the encoding of meaningful gist representations while word concreteness and specific words facilitate verbatim representations. In comparison, FKGL only calculates the number of syllables per word and the number of words per sentence. Therefore, it measures some characteristics of words and sentences but not the relationships between them.

There is growing evidence that GIS predicts knowledge and comprehension of texts about cancer. In behavioral experiments with closely matched pairs of texts about breast cancer, people randomly assigned to read high GIS versions scored significantly higher on a test of breast cancer knowledge, and with a second pair of texts about breast cancer, people randomly assigned to read a high GIS version scored significantly higher on a comprehension test [9]. Another study employed a fill-the-blanks (cloze) procedure to assess comprehension [12,13]. Participants were randomly assigned to read three of nine texts about cancer from the National Cancer Institute (NCI) web pages for patients, including three high GIS, three medium GIS, and three low GIS. As predicted by FTT, applying the cloze procedure to “authentic” NCI cancer texts for patients, GIS predicted subsequent comprehension among randomly assigned readers. An intriguing post hoc finding was that high GIS texts allowed for more word variation retaining a semantically similar meaning for sentences [10]. Unfortunately, an analysis of the complete set of 244 NCI texts about cancer for patients and healthcare providers demonstrated that although the texts for patients were written at a lower readability grade level (as measured by FKGL), the texts for healthcare providers actually scored significantly better on GIS [10]. This suggests that the construction of NCI patient texts is suboptimal in helping everyday readers make meaningful inferences to mentally represent the bottom-line meaning.

It appears that some of what makes high GIS texts more likely to yield inferences is the characteristic of making it easier for readers to represent the meaning of those texts in their own words rather than in the exact (verbatim) words omitted from the text. This explains the superior cloze performance on high GIS texts, since the within-subject design makes it unlikely to be due to reader characteristics. The purpose of the current work is to examine the underlying psychological mechanisms for superior performance with high GIS patient education texts. The first study tests the hypothesis that high GIS texts make it easier to fill-the-blanks with different words conveying the same meaning. It is a partial replication of the Wolfe and colleagues [10] study with a nationally representative sample of older women. The second study examines free recall of two texts about Soft Tissue Sarcoma, one is a high GIS text for healthcare providers, and the other is a low GIS text for patients. It tests the hypothesis that for high GIS texts, readers recall a significantly greater proportion of core decision-relevant propositions, despite being written at an advanced grade level. Collectively, both studies test the proposition that some texts lend themselves more easily to retaining their basic meaning using different words, and this characteristic of texts is captured by Gist Inference Scores.

2. Study 1: cloze comprehension of Cancer information for patients

2.1. Background and aims

Previous research has found that undergraduate participants who read high GIS cancer texts performed significantly better on tests of comprehension [5,10]. This study extends these findings to a nationally representative sample of women over the age of 45 from a Qualtrics panel [14]. Qualtrics panels have been used extensively in social and psychological research with patients [15,16].

The purpose of this target sample is to determine whether previous findings also hold for an older population that has greater cancer risk and are therefore more likely to search the web for cancer information compared to previous college student samples. We investigated the hypothesis that a high GIS text would yield significantly better fill-the-blanks performance using different words than the medium and low GIS text, but lower performance using the exact words in the text. Previous findings showed that high GIS texts seemed constructed in a way that promotes a more flexible use of language to retain the same meaning [10].

2.2. Methods

A Qualtrics panel provided a nationally representative sample of 61 women over the age of 45 (See Table 1 for demographic information) who read three NCI web-based texts for patients with each participant reading a low, medium, and high GIS text. These texts were selected because the low GIS text had a low FKGL score and the high GIS text had a high FKGL score which provides a robust measure of how GIS measures a separate dimension of comprehensibility. The within-subjects design controls for reader characteristics and provides an excellent assessment of the characteristics of patient education texts themselves, as assessed by GIS. As constructed by Wolfe and colleagues [10], each NCI text had every 10th word removed to create a cloze procedure wherein readers had to fill in the blanks with the best fitting word to retain the original meaning. In a review of readability and comprehension instruments used for web-based cancer information, Friedman and Hoffman-Goetz [17] conclude that, "Cloze is a valid and reliable measure of patient comprehension that can be used to pretest written health information . . . though it has a heavy time burden for administrators and participants.” The low GIS text (GIS = -0.80) is a 316 word passage on Myeloproliferative Neoplasms [18].
medium GIS text (GIS = -0.34) is a 434 word passage on Gallbladder Cancer [19]. The high GIS text (GIS = 0.41) is a 579 word passage on Pituitary Tumors [20]. The order of texts was randomized for each participant.

After completing the cloze task for the three NCI texts, participants completed two tests of health literacy. The Short Test of Functional Health Literacy in Adults (STOFHLA) [17] contains two passages about X-Ray Preparation and Medicaid Rights and Responsibilities in which participants must answer fill-in-the-blank questions with four answer options provided per question. The Cancer Health Literacy Test (CHLT-6) [21] is a 6-item multiple choice questionnaire that asks cancer-related questions such as "A biopsy of a tumor is done to... A. Remove it. B. Diagnose it. C. Treat it."

Following the conclusion of the study, participants’ responses on the three cloze tasks were coded as exactly matching the word removed, retaining the same meaning, and not matching. The same coding procedure was used by Wolfe and colleagues [10].

2.3. Results

A repeated measures ANOVA was conducted on cloze performance. As predicted, the high GIS text (M = 10.1 %) had a significantly higher percentage of different but semantically similar responses than the medium (M = 5.7 %) and low (M = 4.4 %) GIS texts, F(2,180) = 9.25, p < 0.001. For exact words only, the medium GIS (M = 33 %) text had a significantly higher percentage of exact words used than the high (M = 20 %) and low (M = 23 %) GIS texts, F(2,180) = 5.18, p = 0.007. For exact words and semantically similar responses combined, there were no significant differences between the high (M = 30 %), medium (M = 39 %), and low (M = 27 %) GIS texts, F(2,180) = 2.46, p = 0.089. The average score on the CHLT-6 was 82.5 % (SD = 26.1) with 77 % of participants obtaining an adequate health literacy score (following criteria [21]). The average score on the STOHLFA was 90.6 % (SD = 13.5) with 93.4 % of participants obtaining an adequate health literacy score (following criteria from [22]). The effect of health literacy and GIS level on cloze performance was examined with an ANOVA. When excluding cloze responses without entries, performance on the CHLT-6 significantly predicted cloze performance, F(2,130) = 4.0, p = 0.048, but the interaction with GIS level was not significant, p = 0.71. This suggests that GIS is useful for use with readers of varying levels of health literacy.

2.4. Summary

As found by Wolfe and colleagues [10], high GIS texts allowed for correct cloze performance using a wider range of words that retain the same basic meaning in authentic patient education texts. Some written materials lend themselves more easily to readers representing them in their own words, and GIS appears to encapsulate that critical aspect of patient education texts. Interestingly, the medium GIS text was best for filling the blanks with the exact words deleted from the original. It may be that low GIS texts promote neither gist nor verbatim representations. The relationship between GIS and verbatim memory should be studied in future investigations. The fill-the-blanks cloze task is one way to assess a reader’s gist representations. Another method is to ask people to read authentic texts about cancer and have them recall what they read in their own words. That is the approach we took in the second experiment.

3. Study 2: free recall of texts about Soft tissue sarcoma

3.1. Aims

The aim of the second study was to further understand how reading texts about cancer that are low and high on GIS relate to readers’ representations of the information. Specifically, the ability to represent the bottom line meaning in one’s own words, and the focus on decision-relevant information. We compared memory of two texts about Soft Tissue Sarcoma from the NCI web site, one written for patients that had a low GIS and the second written for healthcare providers that had a high GIS. This is a rigorous test of GIS since the patient text was written at a lower grade level for a different audience. We predicted that participants who read the high GIS text would recall a greater proportion of propositions that were identified a priori as central to decision making than participants who read the low GIS text. This classification method is related to van Dijk and Kintsch’s (1983) “macropropositions” which involves parsing text into main ideas and tertiary ideas [23].

3.2. Methods

A sample of 198 Miami University undergraduate students were randomly assigned to read either a low or high GIS text about cancer. They were then asked to “Think about the text that you just read. Please write down as much as you can remember from the

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Study 1 demographic information.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Participants</td>
<td>61</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>100 %</td>
</tr>
<tr>
<td>Male</td>
<td>0 %</td>
</tr>
<tr>
<td>Age</td>
<td>M = 58.5 (SD = 7.22)</td>
</tr>
<tr>
<td>Native Language</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>91.8 % (56/61)</td>
</tr>
<tr>
<td>Other</td>
<td>8.2 % (5/61)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>18 % (11/61)</td>
</tr>
<tr>
<td>White</td>
<td>75.4 % (46/61)</td>
</tr>
<tr>
<td>Black</td>
<td>11.5 % (7/61)</td>
</tr>
<tr>
<td>Asian</td>
<td>4.9 % (3/61)</td>
</tr>
<tr>
<td>Native American</td>
<td>1.6 % (1/61)</td>
</tr>
<tr>
<td>Other</td>
<td>6.6 % (4/61)</td>
</tr>
<tr>
<td>Highest Level of Education</td>
<td></td>
</tr>
<tr>
<td>Less Than High School</td>
<td>11.5 %</td>
</tr>
<tr>
<td>High School</td>
<td>21.3 %</td>
</tr>
<tr>
<td>Some College No Degree</td>
<td>26.2 %</td>
</tr>
<tr>
<td>Associate’s Degree</td>
<td>9.8 %</td>
</tr>
<tr>
<td>Bachelor’s Degree</td>
<td>19.7 %</td>
</tr>
<tr>
<td>Master’s Degree</td>
<td>11.5 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Study 2 participant pool demographic information.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Participants</td>
<td>198</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>70.4 %</td>
</tr>
<tr>
<td>Male</td>
<td>29.6 %</td>
</tr>
<tr>
<td>Age</td>
<td>M = 19 (SD = 1.4)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>80.4 %</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1.6 %</td>
</tr>
<tr>
<td>Black</td>
<td>3.2 %</td>
</tr>
<tr>
<td>Asian/Asian American</td>
<td>10.0 %</td>
</tr>
<tr>
<td>Other</td>
<td>4.8 %</td>
</tr>
<tr>
<td>Academic Year</td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>54 %</td>
</tr>
<tr>
<td>Sophomore</td>
<td>27 %</td>
</tr>
<tr>
<td>Junior</td>
<td>11 %</td>
</tr>
<tr>
<td>Senior</td>
<td>13 %</td>
</tr>
</tbody>
</table>
text. Feel free to use your own words but try to write down as much as you remember.” The participant pool was predominantly white (80.4%) and female (70.4%). Table 2 provides further demographic information.

Participants were randomly assigned one of the two text conditions. The high GIS Soft Tissue Sarcoma text [24] was 433 words with a GIS score of 0.704 and a Flesch-Kinkade Grade level of 15.55. The low GIS Soft Tissue Sarcoma text [25] was 466 words long with a GIS of -0.225 and a Flesch-Kinkade Grade level of 10.79. Both selected texts were deconstructed into a list of decision-relevant propositions and other non-critical propositions that were determined before the experiment. Examples of decision-relevant propositions and non-critical propositions can be found in Table 3. Examples of decision-relevant propositions are, “Growth and/or expansion of vascular anomalies can cause clinical problems” and “limited treatment options are available.” An example of a non-critical proposition is, “the ISSVA classification of tumors is based on the WHO classification.” Following data collection, responses from 20 participants (20 high GIS, 20 low GIS) were independently coded by two experimenters and inter-rater reliability was 78%. Experimenters coded participants’ responses for the ratio of key propositions to total propositions, number of total propositions recalled, key propositions recalled (determined a priori), other propositions recalled, and other proposition inferences. Other propositions and other proposition inferences were considered statements that could not be matched to the propositions in the original text.

3.3. Results

A one-way between-subjects ANOVA was used to examine the effect of GIS level on free recall performance. As predicted, a significantly higher proportion of propositions recalled by participants reading the high GIS text expressed the gist of decision-relevant information than for participants in the low GIS condition. After reading the high GIS text, 41.1% of what participants recalled was key decision-relevant information whereas 28.7% of recalled propositions were decision-relevant for those who read the low GIS text, F(1, 172) = 9.82, p = 0.002. However, participants who read the low GIS text recalled significantly more propositions overall, with a mean of 6.48, compared to those who read the high GIS text with a mean of 4.36 propositions recalled, F(1, 193) = 12.61, p < 0.001. Overall, there were no differences in the number of decision-relevant propositions remembered by group, with a mean of 1.76 for the high GIS group and 1.67 for the low GIS group, F < 1.

3.4. Summary

The text for healthcare providers about Soft Tissue Sarcoma was written at a grade level for juniors and seniors in college. Nonetheless, a greater proportion of gist memory for readers concerned decision-relevant information than for the easier patient version. GIS is not capturing total recall or verbatim memory (indeed people remembered more total propositions with the easier low GIS text). Rather, high GIS texts appear to help readers understand and recall the gist of decision-relevant information.

4. Discussion and conclusion

4.1. Discussion

A major obstacle to improving health outcomes and health literacy is communicating complex medical information to nonexperts. The NCI website itself acknowledges this by having both patient-centered and health-provider-centered texts about cancer information. However, adopting the use of GIS in tailoring medical information to patients may offer an improvement over commonly used but superficial text analysis measures such as FKGL. Study 1 shows that with a representative sample of adult women, GIS measures the extent to which readers can use a wider range of words to retain the same bottom-line meaning as a characteristic of high GIS texts themselves. The high GIS texts also had a higher grade level (FKGL), suggesting that the traditional approach of writing at lower grade levels does not always yield the most beneficial results. Study 2 reveals that GIS predicts that a higher proportion of what readers recall from high GIS texts is decision-relevant information, as predicted by Fuzzy-Trace Theory. Taken together, these results provide a meaningful contribution in understanding how cognitive outcomes can be predicted based on solely observable text characteristics, and the superiority in high GIS texts in communicating the bottom-line meaning of medical information.

4.2. Limitations

This research has some limitations that need to be considered and accounted for in future work. In study 1, the nationally representative sample of women provides a meaningful way to generalize previous GIS findings with college students. However, participants from this Qualtrics sample were motivated by payment and are more accustomed to completing opinion surveys rather than difficult cognitive tasks, such as the cloze tasks. This
may have been a contributing factor to some participants leaving many blanks unfilled. A second limitation concerns the study 2 undergraduate student sample. While findings with undergraduate participants is often replicated with participants closer to the target population [26–28], the generalizability of the results is hindered because undergraduates are not the likely audience of information about Soft Tissue Sarcoma and other cancers. However, this is a limitation addressed by the nationally representative sample used in study 1. While women over the age of 45 do not represent all types of patients or at-risk populations, these studies suggest a degree of generalizability. Future research should further examine the interplay between grade level and GIS in patient education materials and compare high and low GIS texts with a sample of newly diagnosed patients.

A third limitation concerns the selection of authentic patient education texts from the NCI website, rather than experimentally-designed texts, in study 1. While this approach offers more ecological validity, it also cedes some experimental control because it cannot account for the impact of different topics (Myeloproliferative Neoplasms, Gallbladder Cancer, Pituitary Tumors) on comprehensibility in addition to GIS level. It is possible that participants performed better on the Pituitary Tumors cloze task because the topic is inherently more understandable than Myeloproliferative Neoplasms. Lastly, the current studies do not establish meaningful cutoffs for GIS. Future studies are needed to determine the optimal and minimal ranges of GIS in predicting positive reader outcomes.

4.3. Conclusion

Understanding the bottom-line meaning is more crucial to later medical decisions than remembering the verbatim facts. Gist Inference Score is a computational tool that provides an automated and user-friendly method to evaluate large scale texts for their ability to convey the bottom-line meaning. The current studies extend previous research on GIS to a nationally representative sample of older women and the free recall of decision-relevant information. We conclude that high GIS texts lend themselves more easily to mentally representing their basic meaning in the reader’s own words.

4.4. Practical implications

These findings provide meaningful information to medical communicators aiming to distill complicated information (i.e. cancer treatments and risk factors) into a digestible format for everyday readers, including older people with higher health risks shown in study 1. Additionally, these findings may encourage medical communicators to use quantitative measures for evaluating the readability of medical text. Compared to GIS, other unidimensional measures of text readability, such as FKGL, may provide a more superficial analysis of text readability. Although our research materials pitted FKGL against GIS, we envision health care providers using both tools in tandem. Although not a direct replacement for quantitative methods of text analysis, GIS requires fewer resources and can be easily adapted by those without psycholinguistic expertise. Additionally, medical professionals can identify the GIS of their written materials by entering them into the freely available Coh-Metrix webtool (cohmetrix.com/tool) and computing the GIS formula from the output – assisted by a spreadsheet found in Supplemental Materials on the journal web site. Additional steps include simply copying and pasting the Coh-Metrix output into Excel. From there, the spreadsheet in the Supplemental Materials will automatically calculate GIS. In short, GIS provides medical communicators with an automated and user-friendly method to evaluate medical texts for their ability to convey the bottom-line meaning.

Contributions

Study design: MD, CRW; article search: MD, CRW; analysis: MD, CRW; drafting of the article: MD, CRW.

Declaration of Competing Interest

None.

Acknowledgments

The authors would like to thank Chase Tirey and Vasanthi Kalindi for their assistance in data collection, data coding, and establishing inter-rater reliability.

Funding

Study 1 was funded by Doris Bergen Center for Human Development, Learning, & Technology, Miami University, Oxford, Ohio. The sponsor was not involved in the study design, data analysis, interpretation of the data, writing the manuscript, or in the decision to submit for publication.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jpec.2020.02.027.

References


